

Clones: oddities or businesses?

The woes of the biotechnology sector have been compounded by technical obstacles for companies seeking to develop cloning technologies. For benefits to materialize, scientific understanding needs fostering by industry and governments.

Poor Dolly. When the sheep was born in 1997, she was a symbol of the potential of the new technology of cloning. Now she stands by as the company involved in her creation, PPL Therapeutics, struggles to show investors that she wasn't just a flash in the pan. Last week, PPL announced that it has cloned genetically modified (GM) pigs that the company said were another step towards the development of pig organs for human transplants. But the pigs' genetic changes will not solve some of the problems that still confront would-be xenotransplanters. PPL's finding was trumpeted in a press release, not in a peer-reviewed journal, leading observers to declare that the announcement was another example of what is becoming a tradition of using press releases to keep stock prices up.

PPL's share price has fallen substantially since the late 1990s, and recent months have been especially bleak for PPL and its competitors who hope to commercialize cloning. The US company Infigen, for example, hoped to make money by selling cloned livestock to farmers. But it has now begun a restructuring campaign and its partner, Pharming of Leiden, the Netherlands, has gone into receivership. Infigen's former head, Michael Bishop, now says he doesn't think that livestock cloning will ever be profitable. The president of ProLinia of Athens, Georgia, said earlier this month that US investors are slow to cough up cash for cloning, but he hoped that a recent report from the National Academy of Sciences would help to ease their worries. And when a government audit released in May raised concerns about the viability of Advanced Cell Technology of Worcester, Massachusetts, its head did not try to hide his financial woes.

Infigen's current head, Walter Simson, says the company has now set its sights on commercializing proteins produced in the milk of

GM cows — the same strategy that PPL is pursuing. PPL says it will spin off its xenotransplantation and stem-cell work by the end of the year. What is Simson's advice for companies who hope to make a living by selling cloned livestock? "I wish them good luck," he says.

Why is a technology that is said to hold so much promise proving so commercially unattractive? US biotech companies say that regulators are taking too long to approve their technologies. But the real problem is a basic lack of information and scientific understanding. The public and the government need more data on how safe these technologies are for both cloned animals and for the people that would benefit from them. If they hope to build public and regulatory confidence, companies must publish more peer-reviewed data to shed light on safety concerns.

Publication could also help to speed scientific progress on the thorny problems that must be solved if cloning and xenotransplantation are to become commercial realities. Cell nuclear reprogramming is still dimly understood, for example. So cloning remains inefficient — and unprofitable. Scientists say they are making progress on this problem, and will eventually be able to solve many of the problems that have so far plagued cloning. But without immediate answers to the outstanding questions, the first batch of commercial cloning companies were probably doomed to fail.

Of course, commercial ups and downs matter less than making cloning work for patients. This is why government rules on research related to cloning really matter. And if Dolly and her GM farm friends are to be anything more than expensive oddities, governments around the world — especially in the research powerhouses such as the United States — must invest significantly in the basic science. ■

Nature Materials arrives

A new forum for diverse disciplines is launched next week.

Textiles and yarns in which are embedded electronic and optical devices. Catalysts that are dependent on the ability to structure crystalline materials at nanometre scales. Optical switches, amplifiers and lasers that exploit semiconductor dots designed to behave more like atoms than bulk solids. Biological and biomedical devices that exploit the behaviour and interfaces between metals and polymers. Magnetic and electronic devices that rely on electron spin. Fuel cells powered by the transformation of hydrocarbons to hydrogen-rich gas. Materials whose properties stem from mimicry of biological structures.

This diversity demonstrates how radically materials research is set to change the world around us. Many of these examples depend on advances that have taken place in recent years in our ability to engineer materials down to atomic scales. All of them depend on the skills that can be brought to bear on materials and devices by physicists, chemists, materials scientists and engineers. An increasing number of them offer opportunities to biologists too.

Small wonder, then, that *Nature* has long published papers in materials research. And given the field's vigour, multidisciplinary

and sheer scale, small wonder, too, that *Nature's* publishers have decided to launch *Nature Materials*, the first issue of which appears on 2 September. Surveys suggest there is room for a journal that spans all aspects of materials research, with the aim of publishing papers of outstanding interest to materials researchers from all disciplines.

Nature's own ambitions in materials research are the same as ever: to publish papers with the deepest or broadest impact on our understanding of materials' properties and behaviour. *Nature* and *Nature Materials* will be editorially independent. However, if *Nature's* editors find that a paper lacks the breadth of impact appropriate for *Nature*, for example, but is still of outstanding interest, they will, with the authors' permission, pass on the file and referees' reports to *Nature Materials'* editors to help save authors time.

Our experience in launching research journals shows that both *Nature* and its siblings flourish, not least in achieving high impact factors — in many cases (including *Nature's*) at the top of their respective categories. *Nature Materials*, its editors say, is already off to a flying start in the breadth and quality of its submissions. Readers can see for themselves at www.nature.com/naturematerials. ■





Pike strike
Alaska's salmon loss
blamed on predator
invasion
p907



Vive la révolution
French biomedical
reforms pledge cash
for partnerships
p908

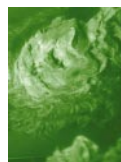


Image conscious
Holograms to reveal
features of 2,000-
year-old bog body
p908



Grape success
Genetic fingerprints
offer way to quash
wine fraud
p909

Central Europe braced for tide of pollution in flood aftermath

Quirin Schiermeier, Munich

The flood waters of the River Elbe in central Europe are in retreat, but a new threat has emerged — dioxins, mercury and bacteria in the sludge carried by the waters.

The Elbe has been receding since 16 August, and sludge clearing is now under way in the Czech Republic and eastern Germany. At the same time, scientists in the region are carefully monitoring the toxins and risk of disease left in the flood's wake.

One major potential source of contamination is Spolana, a chemical plant in Neratovice in the Czech Republic that was submerged until last week. Several disused buildings at the facility are contaminated with dioxins, used in the 1960s to produce a component of the defoliant Agent Orange. The plant, which now makes chlorine-based products such as polyvinyl chloride, also has stores of mercury.

German environment minister Jürgen Trittin and his Czech counterpart Libor Ambrozek say that any toxins in the river would have been diluted below dangerous levels. Preliminary analysis of water from around the plant and at Dresden, 120 kilometres down-river, give no cause for alarm, they say.



Flooding at the Spolana chemical plant has raised concerns over dioxin and mercury contamination.

But Manuel Fernandez, a chemist with the environmental group Greenpeace, says that the buildings used in the 1960s were not properly protected against floods. He adds that Czech police attempted to stop his colleagues from gathering samples near the plant, and that the authorities may try to play

down the true degree of contamination.

But samples taken 300 kilometres down-river at Magdeburg in Germany are being analysed at the German Federal Institute of Hydrology in Koblenz, and initial results show no sign of dioxins. Institute staff are warning that the contamination may not have reached the site when the samples were taken, however.

The tests did find a large increase in levels of arsenic and lead, probably from slag heaps and waste deposits. The Elbe has also flooded hundreds of sewage plants and picked up many animal carcasses. Toxic substances that lay hidden beneath the riverbed may also have been disturbed, says Volkhard Wetzel, head of the Koblenz institute.

The water may also be spreading bacteria such as *Salmonella*. "There is no reason for panic, but we are scrutinizing the situation," says Axel Hofmann, a medical microbiologist at the Saxonian Institute for Health and Veterinary Medicine in Leipzig. Thousands of inhabitants of Prague and Dresden have been vaccinated against hepatitis A, which can be spread by faeces leaking into drinking water.

Dioxin contamination of the Elbe may also hamper efforts to tackle pollution from various industries during the communist era. Ironically, thousands dived into the Elbe in Dresden only last month to mark its recovery from pollution.

Web extends to bottom of the world

David Adam, London

Proof has arrived — if any were needed — that there's no escape from technology. The electronic superhighway is about to reach the ends of the Earth, when the US National Science Foundation (NSF) lays a high-speed fibre-optic cable to the South Pole.

The NSF wants to provide its research base at the pole with continuous electronic communications. It has asked Denver-based Raytheon Polar Services to investigate the practicalities of running a cable across the ice from another, larger research station, over 1,000 miles away.

"This is something we've wanted to do for a while and a feasibility study is being carried out, but it's still very early days," says Peter West, an NSF spokesman.

Polar researchers already have access to

e-mail and the Internet, but only for a few hours a day. The base is the only permanently inhabited place on Earth that lacks a clear line of sight to communications satellites. It relies on intermittent contact with ageing satellites that have drifted from their orbits.

As well as giving 24-hour access to sports results, the link will allow researchers around the world to monitor remote experiments. The South Pole is popular with physicists because the compressed ice is transparent and contains few air bubbles, helping them to spot the faint light emitted from neutrino-particle interactions. Astronomers like it because it offers 'continuous astronomical darkness' — better known as night — for several months of the year, and geologists are interested in the micrometeorites strewn across the Antarctic tundra.

US prepares ground for security clampdown

Erika Check, Washington

The United States is preparing draft guidelines to control the dissemination of a new category of 'sensitive homeland-security information', which could include blueprints of nuclear power plants or research on bioterrorism.

White House officials discussed the plans with representatives from scientific and academic communities on 22 August. But Kathryn Harrington, a spokeswoman for the Office of Science and Technology Policy (OSTP), says that an actual policy will not be released for several months.

Scientists have fretted since September that the government would take draconian action to restrict the publication of scientific information following last autumn's terrorist attacks. Although attendees of last week's meeting say that the administration's ideas on sensitive homeland-security information did not seem overly restrictive, they caution that they have yet to evaluate a specific proposal. Harrington says that this is because the proposal is still being developed by the Office of Management and Budget, at the request of Governor Tom Ridge, director of homeland security. The OSTP is helping the budget office with scientific aspects of the plan.

Some of those who attended the meeting say they are not opposed to the idea of the government taking formal steps to protect certain information. They add that they are pleased that the administration has included



Under wraps: Tom Ridge is demanding fresh security guidelines for sensitive information.

them in early discussions. "It's much better for them and for us to have the chance to discuss things before the plane takes off," says Terry Hartle, a vice-president at the American Council on Education.

But Hartle and others are worried that they left the meeting with little clear idea of just what the administration considers to be sensitive homeland-security information.

"The community is very concerned about what information is going to be involved," says Janet Shoemaker, head of public affairs at the American Society for Microbiology. For

instance, she says, the administration could take steps to protect 'dual-use' research, such as data on deadly pathogens. "There may be good reason to treat some information as a different category," she says, "but we don't want to restrict information that could lead to scientific advancement or medical benefits."

Some meeting participants say they got the impression that the policy will be aimed more at government-owned strategic information — such as plans of possible terrorist targets — than at basic research done in the universities.

Attendees say that government officials seem to be interested in creating a new category of sensitive information primarily because it would be cheaper than fully classifying the information.

Participants also discussed the dilemmas that would accompany a new category of sensitive information, such as how it would be handled at scientific conferences. "Would you deny access to certain people at the meeting, or ask them to leave the room if it came up?" asks Rich Harpel of the National Association of State Universities and Land-Grant Colleges. "It was very unnerving to think about the way this might actually play itself out."

But computer-security expert David Farber of the University of Pennsylvania in Philadelphia — who didn't attend the meeting — says that researchers in his field often sign non-disclosure agreements with industry and that some meetings have long been closed to foreign researchers. ■

Oceanographer navigates path to the Smithsonian

Kendall Powell, Washington

After a couple of years of incessant bad news, researchers at the Smithsonian Institution were expressing some optimism this week following the appointment of a respected oceanographer as the museum complex's undersecretary of science.

Smithsonian secretary Larry Small —

whose business-led approach has drawn him into bitter public conflict with many of the institution's researchers — announced on 19 August that the position will be filled by David Evans, who is currently the senior research administrator at the National Oceanic and Atmospheric Administration.



David Evans plans to provide direction.

Evans has a research background in physical oceanography and its interaction with climate. He succeeds Dennis O'Connor, who left the museum complex in May to become vice-president for research at the University of Maryland at College Park.

Researchers are now hoping that Evans can build a constructive relationship with the different branches of the sprawling Smithsonian empire. Scientists and other critics assailed Small last year for suggesting that the administration of the entire complex should be split in two, with research administered centrally and separately from exhibits. The director of the National Museum of Natural History resigned in protest at the plan (see *Nature* 411, 624; 2001), which is currently being reviewed by the Smithsonian Science Commission, a specially appointed scientific panel.

Evans says that he sees interaction between research and exhibition staff as a unique strength of the Smithsonian — the largest museum complex in the world.

"By using a combined facility of research and museum exposition, we can show people how science is really done," he says. But he asserts that some Smithsonian units have been "left adrift" by the institution, and says that he will draw on his experience to try to provide them with direction.

Institution staff say that they hope Evans's appointment and an ongoing search for a new director for the Natural History museum mean that the Smithsonian is heeding the science commission's call for the prompt hiring of strong scientific leaders. The commission's full report is due by the end of the year.

But some scientists remain sceptical that a central vision for the diverse Smithsonian science units will work. "We do a lot of different things for historical reasons," says Ira Rubinoff, director of the Smithsonian Tropical Research Institute in Panama. "The point is not that we do them all together, but whether it's good science that serves the nation — and I think it does." ■

Fraud inquiry leaves online paper in the ether

Geoff Brumfiel

A nanotechnology paper is spending the summer in a state of electronic limbo, after *Science* published it online but declined to follow up with the usual publication in print.

The article, which shows evidence for the possibility of combining the hot fields of spintronics and molecular transistors, appeared on *Science*'s website Science Express on 18 April. Science Express papers normally appear in the print journal within "several weeks", according to the website. But after the paper's lead author Jan Hendrik Schön, a researcher at Bell Labs in Murray Hill, New Jersey, was accused of fraud in May, the paper's print publication was delayed. A panel chaired by Malcolm Beasley of Stanford University in California is investigating several papers authored by Schön, who has denied the allegations (see *Nature* **417**, 367–368; 2002).

Science would usually send a final version



of the paper, edited for print publication, to the authors for approval, but has not done so in this case, says Eldon Emberly of the NEC Research Institute in Princeton, New Jersey,

a co-author of the paper. *Science*'s editor-in-chief, Donald Kennedy, says that the hold-up is due to the ongoing investigation into Schön's work. Although Kennedy does not know whether this particular paper is part of the investigation, he says that a decision to assign it a page number will be put on hold until the results of an official inquiry are made public later this summer. "If there is a question about a paper, why would we make the financial investment in printing it? That would just be flat-out silly," says Kennedy.

But some other journal editors fear that the situation will confuse researchers, including those who are considering citing the paper, as it will not be clear whether the work was ever published in *Science* or not. Because the paper can be cited by others, it needs to be clear what will happen if it never appears on *Science*'s pages, says Martin Bloom, editor of *Physical Review Letters*, which has considered its online version the "journal of record" since 2001. Bloom adds that his journal would have no choice but to publish an erroneous paper if it had already appeared online. "We would still print it along with a withdrawal," he says, adding that *Science* "needs to set out a policy".

Others worry that the electronic paper, were it to be retracted, might be lost entirely, potentially undermining the integrity of the scientific record. Clear policies are needed for managing online papers, given the fact that they can be altered or removed from websites, says Paul Ginsparg, a professor of physics and information science at Cornell University in Ithaca, New York, and founder of the Los Alamos preprint server — a service that allows scientists to post papers online before peer review. Although *Science* has a right not to publish the paper in print, he says, it ought to leave the original paper online, with appropriate updates on its status.

Nature editor Philip Campbell says that this journal's existing policy would prevent such an occurrence. Whereas Science Express papers pass through a final edit before appearing in print, a *Nature* paper that appears online is considered to be final. Furthermore, authors whose papers are published online are assured that their papers "will be published in print as soon afterwards as space is available". Campbell says that, in a similar set of circumstances, *Nature* would not reconsider its policy.

Kennedy says that any decision about the precise fate of Schön's paper will have to wait until the investigation is published, but that the work will not be removed from the electronic archive. "The paper isn't in limbo at all — it's published," he says. "I really don't think it's a problem."

But Ginsparg disagrees. "The question is: will that URL on Science Express be there 10–20 years from now?" he asks.

Pike pests ravage Alaska's salmon

Rex Dalton, Anchorage

Alaska — whose vast rivers systems might seem pretty resilient to disturbance — is learning a harsh lesson about invasive species.

The northern pike (*Esox lucius*), a fish that already is causing trouble in the rest of the United States, is decimating salmon populations in some rivers in south-central Alaska, researchers in the state believe.

"What was a novelty has turned into a nightmare," says Barry Stratton, a biologist at the State Department of Fish and Game. A growing number of creeks and lakes, once teeming with salmon, are now losing the fish.

Concern about the threat to salmon populations has focused on overfishing and the effect of farmed salmon on natural ones (see *Nature* **416**, 571; 2002). But attention in Alaska is now turning to booming numbers of pike, which eat young salmon in droves before they reach the sea to breed. A 1999 survey by state fisheries managers in Palmer, Alaska, found that 80% of tested pike had salmon in their stomachs.

Alaskan researchers are now applying for money from the federal government to study the impact of the pike and figure out how best to control or eradicate it.

The northern pike occurs naturally in parts of Alaska, but was only introduced to the salmon-rich south-central area in the 1950s, when an angler is thought to have brought it to Bulchitna Lake.

In the late 1980s, flooding spread the pike into the streams of the Susitna and Matanuska river basins, biologists say. Now it has spread through both basins —



On the prowl: the northern pike's spread has crippled salmon populations in Alaskan rivers.

an area the size of Indiana — and elsewhere. For example, the Kenai Peninsula — an ecologically important region of wilderness and parks — now has pike in at least a dozen lakes and four rivers, all of which are rich in salmon and trout.

Stratton says that Alaskan scientists did not initially understand the seriousness of the invasion. "We now know that when a non-native species shows up we have to jump on it," he says.

But some say that it may be too late to save the salmon. "I'm fatalistic about this," says Mark Gamblin, a biologist in Soldotna. "The genie is out of the bottle. It's going to be real difficult to put back the cork."

French agency plans biomedical reformation

Sally Goodman, Paris

Researchers at INSERM, France's biomedical research agency, are being offered more money to establish partnerships with outside universities, hospitals and industry.

At the same time, the agency is trying to reform its relationships with partner organizations inside and outside the government, says Christian Bréchet, INSERM's director. He adds that the reforms may lay the foundations for an entirely new structure for French biomedical research.

This autumn, Bréchet will launch a programme of five-year "interface" contracts for INSERM researchers. These are designed to increase mobility between the agency's labs and external partners. Under the contracts, the researchers will keep their civil-service status and salary, but they will receive an extra 1,500 euros (US\$1,450) a month from one of the partners. In parallel, the partner organizations will run a similar scheme to allow their staff to spend some time in an INSERM lab, with some of the costs of temporarily replacing these staff being covered by INSERM.

"In France, mobility is still often associated with failure rather than career progression," says Bréchet. "That needs to change. This model will combine the security of tenure



Harbinger of change: Christian Bréchet wants to increase mobility among French researchers.

track with the flexibility and stimulation of working with other partners." He hopes that 600–800 of INSERM's 3,000 researchers will be working under the new contracts by 2006.

Bréchet, a hepatitis specialist who took charge of INSERM in February 2001, has already reduced the mandate of the agency's laboratories from 12 to 8 years, with an evaluation every four years. He says that he is now looking at the US National Institutes of Health (NIH) as a model, and wants to start discussions about creating national

"institutes without walls" for cancer, infectious diseases, cardiovascular research, and rare diseases and medical genetics. These would promote more efficient collaboration between researchers from INSERM and other research agencies, he says.

Bréchet's actions could pave the way for a much broader reform of biomedical research in France. The idea of a single, NIH-type agency for life sciences has been talked about in France for more than 20 years. Bréchet believes that the time is ripe for action, and plans to talk to colleagues at other research agencies about the project's feasibility. He says that if the idea is well received, a new agency could see the light of day within four years.

A 1998 attempt to reform INSERM was shelved after opposition from researchers' unions (see *Nature* 395, 630; 1998). This time the unions, although they accept some of Bréchet's arguments, have accused him of not making sufficient use of INSERM's consultative bodies before making decisions.

But Philippe Kourilsky, director of the Pasteur Institute in Paris, which hosts many INSERM researchers, says Bréchet deserves scientists' support. "Researcher mentality has evolved, and the best aren't interested in a job for life any more," he says. ■

Hologram technique to flush out features of bog body

Alison Abbott, Munich

The faces of the ancients are about to be revealed using a toolbox of techniques created by surgeons and engineers.

A team of archaeologists and anthropologists at an interdisciplinary research centre in Germany is using the imaging techniques to reconstruct the features of a 2,000-year-old body.

Found in 1936 in a bog in Lower Saxony, the body's soft tissues have been well

preserved by its stay in the swamp. This is the result of the dehydrating effects of tannic acid, as well as the presence of sphagnum moss in the bog, which hinders decomposition.

The researchers at Caesar, the Center of Advanced European Studies and Research, in Bonn are using computer tomography to get a three-dimensional image of the body's hard tissue, and holography to image the surface of its soft tissues.

Peter Hering, a physicist at the University of Düsseldorf who works at Caesar, first developed the holographic techniques for reconstructing congenitally malformed faces using laser surgery. But in June, he contacted Mamoun Fansa, director of the Oldenburg Nature and Folk Museum, which holds five 'bog bodies' dating back to Roman times. They decided to transfer one of the bodies to Bonn for imaging analysis.

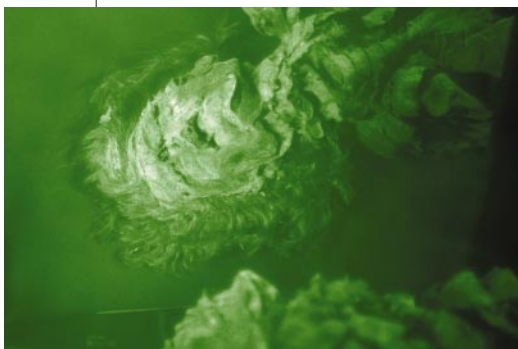
Data from the two analyses are now being merged to create a physical replica of the body, using some prototyping techniques borrowed from the engineering industry. The final reconstruction based on the replica will be crafted by the company Atelier Daynes in Paris, which specializes in

making sculptures of palaeoanthropological specimens.

Reconstructions of ancient bodies based on computer tomography have been done before, but not with the advantage of data on soft tissue, says Hering. "The high-precision sets of data that we have generated will allow the artists to work more accurately and less fancifully," he says.

"Little is known about what humans looked like in the period 300 BC to AD 300 because bodies tended to be burned," says Fansa. Bog bodies are an exception — at least 20 such bodies, thought to be victims of sacrificial rituals, are held by various museums in northern Europe. By studying their faces, historians hope to gain clues as to whether they are Roman or German. Fansa also plans to use DNA analysis to try to resolve this question.

Hering and Fansa are keen to extend the study if the initial analysis works well. Hering says he would like to be able to make holograms of any newly discovered bodies before they are exhumed. The scientists add that the hologram technique will allow other museums to display virtual, life-sized bog bodies. ■



Head start: a holographic image of a bog body's soft tissue will aid its reconstruction.



In the pink: PPL Therapeutics says its pigs lack a sugar that causes human rejection of pig organs.

'Transplant-friendly' pigs stir up debate over peer review

London The birth of five genetically engineered piglets has raised hopes that organ transplants from pigs to humans could become a reality — and started up an argument about the use of peer review.

On 22 August, the Edinburgh-based company PPL Therapeutics announced the birth in July of piglets lacking both copies of a gene involved in pig-to-human organ rejection. The gene produces an enzyme responsible for adding a sugar known as alpha-gal to the surface of pig cells. The sugar triggers the immediate rejection of transplanted pig organs in humans.

The decision to press release the result has annoyed some scientists. "The scientific community will be sceptical about this so-called breakthrough until it has been subject to rigorous peer review," says Patrick Bateson, vice-president of the Royal Society.

PPL attracted similar criticism in January when it announced the birth of pigs in which a single copy of the gene had been knocked out, again without peer review (see *Nature* 415, 103; 2002). That announcement was made the day before another group published similar results.

NASA picks up the pieces of fragmented comet craft

San Francisco CONTOUR, NASA's ill-fated comet-hunting spacecraft, has broken into at least three pieces, telescope images from observatories in Arizona, California and Hawaii have confirmed.

The craft's launch on 3 July went as planned, but it broke up after an onboard rocket booster was fired on 15 August (see *Nature* 418, 806; 2002). The rocket should have propelled CONTOUR out of its Earth orbit and towards a rendezvous with Comet Enke in 2003. The observed fragments are travelling on CONTOUR's predicted trajectory but are now so far away that further observations are unlikely.

NASA officials say that CONTOUR may still be capable of operating, and intend to continue monitoring until early December,

when the spacecraft will come into a more favourable orientation for receiving signals from Earth.

Neutrino detector to recover from shock

Tokyo The Japanese government is set to approve a four-year plan to restore the Super-Kamiokande (Super-K) neutrino detector. Officials hope to have the detector back to full capacity by 2006.

Super-K uses around 11,000 light-sensitive detectors to spot light released from collisions between neutrinos and water molecules in its 50-million-litre tank. More than half were destroyed when a fault caused a shock wave to rip through the tank last November.

Some physicists have been concerned that the government might be reluctant to repair the detector at a time when many big Japanese science projects are being cut (see *Nature* 416, 118–119; 2002), but officials at the education ministry say that letters from high-profile overseas scientists helped to convince them of the need for refurbishment. The ministry has included the money needed for the first 1,000 replacement detectors in its budget request for the next financial year.

Super-K should begin taking readings in December using the sensors that were not damaged in the accident, but the project will restore the sensitivity needed for other studies, such as those of solar neutrinos.

DNA test could end the party for shady wine dealers

Paris Out with the gargle-and-spit approach; in with genetic fingerprinting. French researchers say that DNA analysis could be used to identify the grapes that have been used in a particular wine, and hence to combat wine and vine fraud.



Makes you spit: quality producers currently lose out to wine fraud.

A team at the Montpellier research centre of INRA, the French agronomic research institute, has found a way to trace molecular markers of individual grape varieties in any vine product that contains sufficient DNA. The technique has already been successfully used to trace the origin of grapes for eating. The researchers are now turning their attention to commercial wines, and are refining

their technique in the hope that it can eventually be used for purified wines, which contain extremely low levels of DNA.

France's anti-fraud agency carries out thousands of investigations each year to try to stamp out wine fraud. In May, a French wine dealer was given an 18-month prison sentence and a fine of 1.25 million euros (US\$1.2 million) for selling Bordeaux cut with a cheap table wine under a good-quality label.

Team to decipher frog genome hops to it

San Francisco The genome sequence of the African frog *Xenopus tropicalis* will soon be available, courtesy of researchers at the US Department of Energy (DOE)'s Joint Genome Institute (JGI) in Walnut Creek, California. The team announced on 20 August that it will receive funding from the DOE, effectively giving the project the green light. The researchers hope to publish a draft sequence by the end of 2004 and a full version, in which each base will have been checked eight times, about six months later.

X. tropicalis was chosen, rather than the more commonly used *X. laevis*, because it has a much smaller genome and shorter life cycle. Researchers use *Xenopus* to study cell and organ development, including fundamental processes such as cell communication and division. The JGI team says that the genome sequence will shed light on how these can go wrong, providing insights into human health issues.

♦ www.jgi.doe.gov

Word wizard creates a font of knowledge for physicists

Washington High-energy physicists may be able to create exotic subatomic particles, but it took an English professor to help them conquer Microsoft Word.

University of Mississippi physicist James Reidy was frustrated at the cumbersome software needed to generate symbols for antimatter particles, which have a line over the corresponding matter symbol. He teamed up with campus Macintosh guru T. J. Ray of the English department to create LinguistA, a new font that allows easy overscoring of letters, numbers and Greek symbols when using Word on a Mac. Users simply press shift-5 and type in the 'normal' character to generate an overscore.

"It has saved me an awful lot of time," says Reidy. He posted the font on the Cornell University Library arXiv e-print server "as a lark", and was surprised at the response. Overscored symbols can, for example, also be used in biological equations and to denote chemical bonding. PC users will have to devise their own font, however, as Reidy says his department's loyalty lies with Macs.

♦ www.phy.olemiss.edu/HEP/LinguistA.sea.hqx

Almost human...

Sequencing the chimpanzee has emerged as a top genomic priority. David Cyranoski asks the chimp's champions what they hope to gain from studying the genome of our closest living relative.

It was a shocking disappointment to some, and a pleasant surprise to others. On 22 May, when the US National Human Genome Research Institute (NHGRI) announced its list of top-priority organisms for genome sequencing, the chimpanzee had made the cut¹. The rhesus macaque, the most widely used primate in biomedical research, had not.

Arguments about the merits of the decision are still rumbling on. But now the chimp has received the nod, what scientific advances can we hope for? The common chimpanzee (*Pan troglodytes*) is our closest relative. Its genome sequence is about 98.8% identical to our own², and we shared a common ancestor some six million years ago. So one major hope is that the differences between the sequences will reveal the genetic basis for our mental and linguistic capacities, and explain why we are susceptible to some diseases that do not afflict the great apes.

"It will tell us what makes us human," claims Yoshiyuki Sakaki of the University of Tokyo's Institute of Medical Science, who is president of the Human Genome Organization and part of a Japanese consortium that is already sequencing the chimp's chromosome 22 (see 'Will Japan lead the way?', overleaf).

Despite these grand goals, some advocates of chimp sequencing admit that they were surprised by the NHGRI's decision. Given that the rhesus macaque (*Macaca mulatta*) is an important experimental model in fields from cognitive neuroscience to HIV vaccine research, they had expected it to edge the chimp out of the list of top sequencing priorities — but they are delighted with the result. "It was a choice for philosophy and evolution, against current trends towards biomedical

application," says Svante Pääbo, a geneticist at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, who wrote a 'white paper' urging the NHGRI to support chimp sequencing.

But other chimp advocates argue that the animal's genome sequence will have important biomedical applications. "Our case was overwhelming," concludes Maynard Olson of the University of Washington in Seattle, co-author of another pro-chimp white paper submitted to the NHGRI.

Spot the difference

With complete genome sequences from a variety of model organisms now finished or nearing completion, comparative genomics is in full swing. But the chimp will bring a new twist. Until now, researchers have mostly looked for regions of similarity between the genomes of relatively distantly related organisms — reasoning that these must have been conserved by evolution because they have an important function. "Comparative genetics has, until today, looked for conserved regions, but with the chimp we will be looking for differences," says Pääbo.

In the first instance, about 90% of the differences that are uncovered will represent sequencing errors, says Evan Eichler of Case Western Reserve University in Cleveland, Ohio. But although pinpointing such errors will help with the 'proofreading' of the human genome, genuine differences are what really fascinate researchers such as Eichler.

Documenting the differences between the human and chimp genomes will let researchers tackle some thorny problems concerning evolutionary change. A recent study comparing humans, rats and mice³, for

example, suggested that the rate of sequence divergence in mammals has been different for different chromosomes. Preliminary comparisons between the human genome and a sample of chimp DNA sequences have reinforced this view⁴.

To conduct such studies, researchers line up genome sequences next to each other and pick out the differences. This is easiest to do for protein-coding genes — these can be rendered useless by even small changes, so they are relatively well conserved even among distant species. But non-coding regions are harder to align, especially when the structure of the chromosomes on which they sit has drastically diverged — as is the case for rodent–human comparisons. "The problem is knowing whether apparent variation in rates of evolution is real, or due to alignment artefacts," says Laurence Hurst, an evolutionary geneticist at the University of Bath, UK. A complete chimp genome, which is structurally similar to our own, would help to resolve this problem.



Primate objective: Maynard Olson wants to see what genes chimps have that we don't.

NATIONAL GEOGRAPHIC/GETTY IMAGES

In addition, human–chimp comparisons should shed light on ‘hypervariable’ regions of the human genome. In such regions, the rate of mutation or large-scale chromosomal rearrangements is so high that it is impossible to make any sense of comparisons with the mouse, with which we last shared a common ancestor some 60 million years ago. “The chimpanzee sequence will allow us to reconstruct the history of events and therefore infer the mechanisms involved,” says Eichler.

Meet the ancestors

Genome-wide studies of single-nucleotide polymorphisms (SNPs) could also gain from comparisons with the chimp. SNPs are points in a genome sequence where different individuals can have a different base, or letter of the genetic code. They account for most of the genetic variation across human populations, and can be used to map the locations of genes that influence susceptibility to disease. The chimp genome will help researchers pick out the ancestral form of the variation — assumed to be the one shared with the chimp — and separate it from those derived after the two species diverged.

Leonid Kruglyak of the Fred Hutchinson Cancer Research Center in Seattle, who develops computational models to help map the location of disease genes using SNPs, says that knowing which form is ancestral can help infer the history of human populations, including ancient migrations. This information may help researchers to find genes associated with common conditions such as cancer and cardiovascular disease, he argues.

For many biologists, human–chimp comparisons will prove of greatest value in revealing the molecular secrets of human language and cognitive abilities. This is already moving forward following the discovery of a gene called *FOXP2*, mutations in which cause a severe speech and language disorder⁵. Pääbo and Tony Monaco of the University of Oxford, UK, have now shown that the human version of the gene produces a protein that differs from that of the chimp at two amino acids⁶.

Advocates of chimp sequencing also hope to test what Olson calls the “less is more” theory of human evolution⁷. This argues that many of the genetic changes that separate humans from chimps, and explain our success in a wide variety of environments, involve the loss of genes that adapted primates to their forest habitats. Our decreased muscular strength, loss of body hair and delayed maturation all fit with this idea, as does the observation that there appears to be less genetic diversity in the entire human species than in the small remnant chimp populations of central Africa⁸.

At present, there are few data to test the theory at the level of individual genes. But researchers led by Ajit Varki of the University of California, San Diego, found in 1998 that humans have lost an enzyme that adds a



Comparison of disease susceptibility between chimps and humans

Condition	Human	Chimp
Definite differences		
HIV progression to AIDS	Common	Very rare
Influenza A symptoms	Moderate/severe	Mild
Hepatitis B/C late complications	Moderate/severe	Mild
<i>Plasmodium falciparum</i> malaria	Susceptible	Resistant
Menopause	Universal	Rare
Probable differences		
<i>Escherichia coli</i> K99 gastroenteritis	Resistant	Sensitive?
Alzheimer's disease pathology	Complete	Incomplete
Coronary atherosclerosis	Common	Uncommon
Epithelial cancers	Common	Rare

Source: Olson, M. V. et al. A white paper advocating complete sequencing of the genome of the common chimpanzee, *Pan troglodytes*. (2002).

molecule called Neu5Gc, a type of acidic sugar known as a sialic acid, to proteins carried on cell surfaces⁹. More recently, Varki's team has shown that we also lack a cell-surface receptor that recognizes sialic acids¹⁰. Both are present in chimps.

Missing links

Might such genetic differences explain differences in disease susceptibility between humans and chimps? Olson, for one, believes that genetic loss in the human lineage could explain why we are plagued by certain viruses, cancers and degenerative diseases that pose chimps few problems (see table, above). The sialic-acid findings are intriguing because differences in cell-surface molecules might, for instance, influence the ability of certain viruses to infect human, as opposed to chimp cells.

In particular, the two species' different susceptibility to HIV holds the compelling prospect of revealing the genetic changes that have made the virus such a scourge of human health. “Chimps are infected with the

very same virus that infects humans, and very rarely progress to AIDS,” observes Varki. One company, Evolutionary Genomics of Aurora, Colorado, is already looking for differences between human and chimp genes that might explain this difference. “We want to know how one species has already solved the AIDS question,” says Walter Messier, its chief scientist.

Intriguingly, researchers led by Ronald Bontrop of the Biomedical Primate Research Centre in Rijswijk, the Netherlands, argued earlier this month that the key may lie in major histocompatibility complex (MHC) genes, which are central to immune responses. Chimps appear to have a less diverse MHC than humans, suggesting that a viral epidemic in their evolutionary history selected against individuals carrying MHC variants that conferred susceptibility to HIV-like viruses¹¹.

But it is possible that studies of protein-coding gene sequences may leave the difference between humans and chimps shrouded in mystery. Only a handful of genes that differ between human and chimp have so far been

found¹². And many experts suspect that there won't be a huge number more. "I would be very surprised if there are more than a few," says Tetsuo Yamamori of the National Institute for Basic Biology near Nagoya, Japan.

Instead, the differences between chimp and human might largely be the result of gene-regulatory mechanisms giving different levels of gene expression. Pääbo's team has already used DNA microarrays and gel electrophoresis to investigate differences in gene expression and protein production between humans and chimps¹³. The researchers found that these differences were much more pronounced in the brain than in the liver or blood — which opens new avenues for evolutionary studies of human brain development.

But understanding why expression levels differ between humans and chimps may be difficult. One obvious place to look is at the promoter and other regulatory sequences that sit alongside genes and drive their expression. But the control could also be exercised on more subtle levels, for instance through the operation of non-coding RNAs transcribed elsewhere in the genome — some scientists speculate that these may work in networks of gene regulation by interacting with DNA, RNA and proteins¹⁴.

Rhesus positive

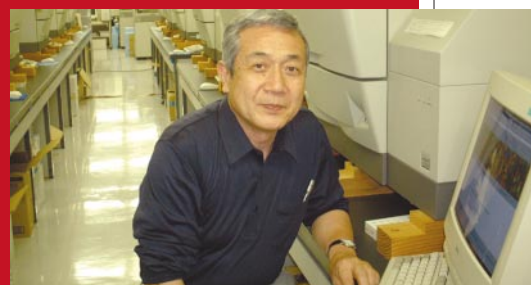
While supporters of chimp sequencing speculate about the findings that lie ahead, the rhesus macaque's supporters are still fuming, and remain convinced that their animal should have been chosen. "Everyone I know — everyone — in my circle of friends in the primate-research community feels unquestionably that the rhesus monkey is the best choice," says virologist Ronald Desrosiers of Harvard University, director of the New England National Primate Research Center in Southborough, Massachusetts.

The main reason for wanting to sequence the rhesus macaque's genome is that it would aid interpretation of the results from the wide variety of experiments in which the

Will Japan lead the way?

Although several groups have sequenced scattered fragments of the chimp's genome, the largest systematic effort has been in Japan. By the end of this year, a Japanese-led consortium hopes to finish sequencing chimp chromosome 22. This effort has drawn heavily on the experience of Yoshiyuki Sakaki of the University of Tokyo's Institute of Medical Science, who led Japan's contribution to sequencing the homologous human chromosome 21.

Sakaki and Naruya Saitou of the National Institute of Genetics in Mishima, south of Tokyo, have repeatedly applied for grants to expand the project, but have so far been unsuccessful — despite strong lobbying from



Off and running: Yoshiyuki Sakaki is helping to sequence chimp chromosome 22.

Akiyoshi Wada, director of the Genomic Sciences Center in Yokohama, where Sakaki has a joint appointment.

"Japan is two years ahead in sequencing the chimp, and so it should take the lead with the project,"



Akiyoshi Wada wants to keep Japan's lead.

says Wada, who complains that Japan always seems to await a cue from the United States before committing to large projects. "Only then will Japan move, but then it's already too late," sighs Wada, who feels a sense of *déjà vu* — he urged Japan to begin high-throughput sequencing of the human genome in the late 1980s, before the project built momentum elsewhere.

The chimp project also faces opposition from the leaders of rival genomic projects.

One floor up from Sakaki at the Institute of Medical Science is the lab of Yusuuke Nakamura, director of the Human Genome Center, who is genotyping vast numbers of human single-nucleotide polymorphisms for use in searching for disease susceptibility genes. He has been an outspoken critic of the chimp project, arguing that taxpayers' money should go to projects with more obvious medical benefits.

Japanese supporters of chimp sequencing feel vindicated by the US National Human Genome Research Institute's decision to make the chimp a top sequencing priority. "Of course there is also a competitive aspect to it," admits Sakaki, who hopes that US interest will spur the Japanese government to provide further support.

species is involved. AIDS researchers, who study a virus called SIV in macaques that causes a disease very similar to human AIDS, are particularly disappointed — and see a bitter irony in the arguments being made for the chimp's value to HIV research. "Most research on AIDS now is on preventing progression of the disease," says John VandeBerg, director of the Southwest National Primate Research Center in San Antonio, Texas. For this, he argues, the rhesus macaque is by far superior. Desrosiers agrees: "Experimental infection of rhesus macaques is the only model on the radar screen for AIDS."

Some researchers even question the chimp advocates' strongest suit — that of evolution. They claim that the macaque's genome sequence, which differs from ours by some 5–7%, will be more informative for many comparative studies.

Then there is the fact that, for ethical reasons, chimps cannot be widely used in invasive biomedical experiments — or even be kept under the uniform lab conditions that may be necessary to produce meaningful results. "Patterns of gene expression in different areas of the brain will vary so much just due to environmental factors that it is often impossible to tell human patterns apart from chimp patterns," says Yamamori. "The macaque, whose activity, diet and lifespan can be regulated and kept constant by researchers, thus offers a better research tool."

NHGRI officials have answered these criticisms by noting that there is no firm commitment to sequencing any organism — all that has been produced is a list of priorities to be considered as spare capacity opens up at the sequencing centres now finishing the human genome. They say that the list is not fixed, so the macaque may get included at a later date.

Ideally, say researchers on both sides of the debate, both chimp and macaque would be sequenced, along with a range of other primates to allow extensive comparative studies. "With advances in sequencing capacity, I hope this unfortunate competition will become a moot point in a few years' time," says Varki. ■

David Cyranoski is Nature's Asian-Pacific correspondent.

1. Check, E. *Nature* **417**, 473–474 (2002).
2. Fujiyama, A. *et al. Science* **295**, 131–134 (2002).
3. Lercher, M. J., Williams, E. J. B. & Hurst, L. D. *Mol. Biol. Evol.* **18**, 2032–2039 (2001).
4. Ebersberger, I., Metzler, D., Schwarz, C. & Pääbo, S. *Am. J. Hum. Genet.* **70**, 1490–1497 (2002).
5. Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F. & Monaco, A. P. *Nature* **413**, 519–523 (2001).
6. Enard, W. *et al. Nature* advance online publication, 14 August 2002 (doi:10.1038/nature01025).
7. Olson, M. V. *Am. J. Hum. Genet.* **64**, 18–23 (1999).
8. Kaessmann, H., Wiebe, V., Weiss, G. & Pääbo, S. *Nature Genet.* **27**, 155–156 (2001).
9. Chou, H.-H. *et al. Proc. Natl Acad. Sci. USA* **95**, 11751–11756 (1998).
10. Angata, T., Varki, N. M. & Varki, A. *J. Biol. Chem.* **276**, 40282–40287 (2001).
11. de Groot, N. G. *et al. Proc. Natl Acad. Sci. USA* advance online publication, 19 August 2002 (doi:10.1073/pnas.182420799).
12. Gagneux, P. & Varki, A. *Mol. Phylogenet. Evol.* **18**, 2–13 (2001).
13. Enard, W. *et al. Science* **296**, 340–343 (2002).
14. Dennis, C. *Nature* **418**, 122–124 (2002).



Talking point: Svante Pääbo believes the chimp genome will shed light on our linguistic abilities.

Fire from ice

Natural gas is in great demand, and researchers know where vast amounts are hidden — in icy crystals called hydrates. But getting it out is another matter, as David Adam finds out.

S. DALLIMORE/GEOL. SURVEY CAN.



T. COLLETT/US GEOL. SURVEY



What a gas: researchers at Mallik Field, Canada, used a drilling rig (top) to mine methane from buried hydrate crystals (above).



R. LAFRAMBOISE/GEOL. SURVEY CAN.

It's unlike fossil-fuel companies to ignore an untapped energy resource. Yet from the Arctic tundra to the Indian Ocean, huge reserves of methane lie undisturbed, trapped in strange, ice-like crystals. Estimates of the extent of these deposits vary wildly, but they could contain more than twice as much methane as known conventional gas reserves¹. If just a fraction of this could be recovered, the deposits would be a viable energy source.

The energy companies' reticence is understandable, however. Extracting methane from the crystals, which are known as hydrates, is costly. The deposits are also considered a nuisance by many. Catastrophic releases of gases from hydrates disrupt attempts to drill for oil, and a few researchers have speculated that they could lie behind the mysterious disappearance of some ships.

But things could be changing. The US National Petroleum Council, a government advisory body, estimates that shifting from coal- and oil-fuelled power stations to cleaner, methane-powered alternatives, together with increased use of natural gas in homes and industry, will push up US demand for methane by 40% over the next 15 years. And as the importance of secure energy sources grows, prompted in part by continued instability in the Middle East, so does the need for more hydrates research.

The deposits form when high pressure causes water in sediments to freeze at higher temperatures than normal, giving rise to a solid containing unstable cavities rather

than typical ice. As the temperature continues to fall, the cavities collapse, forming normal ice. But in some sediments, bubbles of methane rise up from underlying reservoirs and fill the cavities, preventing them from collapsing. The result — methane hydrate — looks like normal ice but burns if touched by a flame.

Fuelling progress

In the past decade, several countries have set up research programmes focused on hydrates — Japan, for example, hopes that hydrates could contribute to its energy production within a couple of decades. Estimates of the amount of gas that could be extracted are encouraging, and tentative drilling experiments have been successful. Earlier this year, researchers successfully tapped subterranean deposits for the first time. "As long as energy supplies are not threatened, people can always find cheaper fuel than hydrates," says Bahman Tohidi, head of the Centre for Gas Hydrate Research at Heriot-Watt University in Edinburgh, UK. "But now the political situation is not so stable, methane hydrates begin to look more attractive."

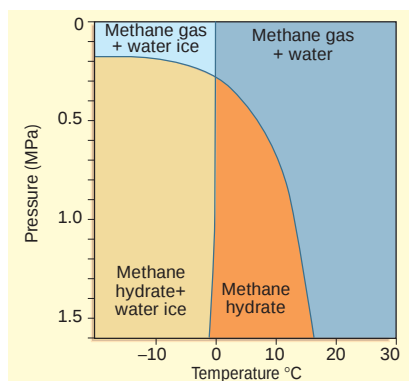
Hydrates form only within a narrow range of pressure and temperature (see chart, overleaf), so they are restricted to certain locations. Some deposits are hundreds of

metres beneath the sea floor, in areas where the water is more than half a kilometre deep. Similar conditions cause hydrates to build up around a kilometre below the Arctic tundra. But although geologists now find it relatively easy to locate hydrate deposits, extracting the methane is another matter.

Additives such as methanol can be used to release the methane in a controlled manner, but most experts fear that the chemicals could, in some locations, pollute nearby aquifers. Reversing the conditions that led to the hydrates' formation is one alternative. Most extraction schemes involve heating hydrate deposits, reducing the pressure on them, or a combination of the two — but each technique comes with problems.

Using hot water to release the gas requires large amounts of energy. Reducing the pressure on the deposits also releases methane but, although less energy-intensive, this method cools the surroundings, perhaps enough to make the hydrates reform. Brad Tomer, head of the US Department of Energy's (DOE) National Methane Hydrates Program at the Strategic Center for Natural Gas in Morgantown, West Virginia, thinks this could be avoided by continuing to reduce the pressure, but admits that long-term tests would be needed to evaluate this idea.

Attempts to compare these methods began last January. Most usable hydrate



Phase chart for a mixture of methane and water; pressure increases with subterranean depth.

deposits probably lie offshore, but it is cheaper to begin with those beneath the Arctic. One of the most promising sites lies in Mallik Field in Canada's Northwest Territories. As well as abundant hydrates, the site has similar geology and reservoir conditions to many offshore deposits, making it an ideal and accessible testing ground.

Last winter, a group of researchers from Canada, Japan, India, Germany and the United States began probing the hydrates. The bulk of the researchers' experiment, led by the Geological Survey of Canada, involved drilling into the hydrates, circulating warm water for several days and measuring the amount of gas produced. They also carried out pressure-reduction trials by carefully sucking out residual water.

Turn on the flare

Those involved say that they were encouraged by the amount of methane produced — enough to ignite a flare similar to those seen burning over oil rigs. But whether the yellow flame is symbolic or a genuine step forwards remains to be seen. The researchers are confident that they will get more energy out than they put in, but that alone will not make extraction economically worthwhile. It might be financially viable to power an on-site turbine, for example, but not to pipe it to a power station. Exact details of the results are being kept under wraps until 2004, when confidentiality agreements with energy companies that helped to fund the work expire.



Low risk: Timothy Collett (far left) is confident that hydrate mining is not unduly dangerous.

DOE-funded researchers will begin examining a different — and potentially more promising — production technique this winter in Alaska. Many hydrate deposits lie above conventional methane reservoirs. Draining this gas could reduce the pressure on the hydrates, causing them to melt and allowing the escaping gas to top up the reservoir. "It's likely that the first production from gas hydrates will be made economically viable by free gas below," says Timothy Collett, a hydrates researcher with the US Geological Survey in Denver, Colorado.

The Indian and Japanese governments agree — both are attempting to produce methane from offshore deposits in this way. In Japan's case, the focus is the Nankai Trough off the country's southern coast. Tetsuo Yonezawa, who heads the hydrates programme of the Japan National Oil Corporation's Technology Research Centre in Makuhari, hopes to begin commercial extraction within 15 years.

No one has proved that this method will work, but a natural methane reservoir topped up by dissociating hydrates may already exist. Methane was extracted from the Messoyakha gas field in Siberia from 1969 to 1985, but the pressure drop in the field was less marked than expected. A layer of methane hydrate is thought to overlie part of the reservoir, and many experts argue that it is producing replacement gas and so helping to maintain the pressure. Not everyone agrees, however. Methane reservoirs do not release gas at constant pressure, and others argue that the pressure change is within normal variability².

The issue might become clearer if researchers can find out exactly how big the Messoyakha deposits are, but hydrates are hard to identify remotely. Engineers can use seismic waves to identify the presence of hydrates, but cannot deduce how much methane they contain. Drilling for samples can confirm whether hydrates actually exist, but the deposits occur both as relatively large layers and as smaller nuggets, so it is difficult to build up a big picture in this way.

By a landslide

This makes it difficult to know how much gas can be exploited by draining an underlying reservoir. But the uncertainty also gives researchers other worries. Hydrates could lend mechanical strength to the sediment. Removing them might make the sea floor fail, producing underwater landslides and enormous uncontrolled gas releases. Massive releases of methane in the geological record are, for example, associated with the collapse of seafloor slopes³. Right now, this is a worrying possibility that hydrates researchers simply have to live with. "You always keep in mind that the extreme is possible," says Collett. "But the likelihood of it happening is pretty low. After all, we probably drill into hydrates somewhere in the world every day."

The possibility of such events is behind



Brad Tomer is optimistic that hydrates can be tapped by reducing the pressure on the deposits.

one of the most enduring stories about hydrates: that sudden releases of gas could sink ships by making the water below the vessel less dense. Some researchers have speculated that an escape of methane could have sunk the *Gaul*, a trawler that went down in the Barents Sea in 1974. The loss of the *Gaul* attracted attention because no distress call was ever received, and the wreck, which was discovered in 1997, appears to be in good condition.

Such claims are greeted with scepticism by most hydrates researchers. In the case of the *Gaul*, an investigation by the British Marine Accident Investigation Branch, released in 1999, suggested that high waves had caused the vessel to capsize, although the issue is being further investigated by a government inquiry. But engineers say that gas releases can sink ships — although it has nothing to do with buoyancy.

Energy companies, for example, have lost ships when large amounts of methane have been released from conventional reservoirs during drilling projects. The force of the escaping gas tips the boat, causing it to capsize and fill with water through open hatches. "Poor seamanship, together with a very large hydrate release, could provide conditions conducive to sinking," says Jerome Milgram, an ocean-engineering researcher at the Massachusetts Institute of Technology who has studied such accidents.

Only time will tell whether the uncertainties surrounding the extraction of methane from hydrates can be cleared up. Even the most optimistic advocates concede that lack of knowledge about the distribution of hydrates, and the methods needed to extract them, present large obstacles. But many researchers remain hopeful. They point out that other commercial energy sources, such as the pockets of methane gas found alongside coal, initially seemed impossible to exploit. Necessity is the mother of invention, after all. If the need for natural gas continues to increase, some of the problems will undoubtedly seem less daunting. ■

David Adam is a news and features writer for Nature.

1. Kvenvolden, K. A. in *The Future of Energy Gases* (US Geol. Survey Professional Paper 1570) (ed. Howell, D.G.), 555–561 (US Geol. Survey, Reston, Virginia, 1993).
2. Collett, T. S. & Ginsburg, G. D. *Int. J. Offshore Polar Eng.* **8**, 22–29 (1998).
3. Dickens, G. R., O'Neil, R. R., Rea, D. K. & Owen, R. M. *Paleoceanography* **10**, 965–971 (1995).

US National Methane Hydrate Program
www.netl.doe.gov/scng/hydrate

Vaccines should be kept even if polio is wiped out

Sir — We would like to clarify a point in your News story (*Nature* **418**, 265; 2002) about our report of the chemical and biochemical synthesis of poliovirus (J. Cello, A. V. Paul & E. Wimmer, *Science* **297**, 1016–1018; 2002). The World Health Organization (WHO) is considering terminating polio vaccination after global eradication has been certified, when wild-type polioviruses would no longer be circulating in the human population. But the WHO is also formulating plans to stockpile vaccines in case an outbreak occurs after termination, caused by viruses that had been undetected during the eradication campaign. We suggest stockpiles should be maintained indefinitely if vaccination against polio is terminated.

Jeronimo Cello, Aniko V. Paul, Eckard Wimmer

Department of Molecular Genetics and Microbiology, School of Medicine, State University of New York at Stony Brook, Stony Brook, New York 11794-5222, USA

Arp2/3 plays no role in muscle contraction

Sir — We are pleased that Arp2/3 complex was included as an example in your News Feature “The society of proteins” (*Nature* **417**, 894–896; 2002). However, to the uninitiated reader it might appear that Arp2/3 complex is involved in muscle contraction. Although it participates in the assembly of actin filaments at the leading edge of motile cells, we want to clarify that it has no known role in muscle in either the assembly of actin filaments or contraction.

Thomas D. Pollard*, Dorit Hanein†, Niels Volkmann†

*Yale University, KBT 548, New Haven, Connecticut 06520, USA

†Burnham Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, USA

Seeing stars

Sir — Your Opinion article “Save starry nights” (*Nature* **418**, 709; 2002) advocates control of public lighting to prevent a lack of starlight diminishing the quality of city-dwellers' lives. As you point out, someone in Venice can look up and see all the stars in the constellation of Ursa Minor, but this would be impossible for people in most other cities from New York to Sydney.

Reduction of light pollution may, indeed, help New Yorkers to see these stars. People in Sydney, however, will still have to

go to a planetarium to see Ursa Minor, as they live in the Southern Hemisphere.

Rob A. Gruters

UMR2142 CNRS/bioMerieux, ENS de Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France

Thriving UCSF needs all the space it can get

Sir — You state in a News story (*Nature* **418**, 355; 2002) that “biotech firms are turning down the chance to move into premises on a top university campus”. The concept that this university site is unattractive to companies is central to this article and is implied by the headline.

The University of California, San Francisco, (UCSF) is not seeking companies to move onto our property. We need every square foot we can get for our own laboratories at our new 43-acre

research and teaching campus at Mission Bay. There is no \$85-million commercial site vacant on UCSF property as stated in your article, nor will there ever be.

Your article confuses the Mission Bay campus with an adjacent 300-acre site owned by Catellus, a real-estate developer. UCSF's site is proceeding spectacularly, with buildings worth more than \$800 million either under construction or in the planning stages. Attracting companies to the Catellus site is a marketing issue for the privately owned biotechnology park.

Regis Kelly

Executive Vice Chancellor, UCSF, 513 Parnassus Avenue, Room S-101, San Francisco, California 94143-0407, USA

This story was edited for space reasons, cutting out a statement that UCSF's development and the adjacent commercial one are being managed separately — News Editor, *Nature*.

Efforts to revive Serbian science

Members of the 'old guard' remain in power and are hindering moves to change a demoralized culture.

Sir — Since the overthrow of Slobodan Milosevic's regime, many efforts have been made to revitalize Serbia's ruined society. Milosevic's rule was particularly detrimental to science, which was devastated both in manpower and in financial resources. Most top scientists have left the country, and investment in basic research has been reduced to the most dangerous level yet (for example, see *Nature* **394**, 715; 1998 and *Europhys. News* **30**, 4; 1999). What is the standing of fundamental research in Serbia now?

The new Ministry for Science, Technology and Development has been trying to remedy the misdeeds of the previous regime, but the resistance of the 'old guard' is still very powerful. Conflict between the ministry and scientific institutions has resulted in disagreement about the percentage of the gross national product to be allocated to the sciences. One of the latest proposals is 0.21%, a decrease of 37% for fundamental science research. The dilemma is whether a poor country like Serbia should try to reach the developed countries' goal of 1%, which, according to our basic-research institutions, is necessary to perform internationally competitive work. Or should it concentrate on applied research, as the new minister of science advocates?

Another important issue for basic science, in the context of very limited resources, is whether to continue doing

research at specialized institutes (the practice until now) or move it to universities, as in the United Kingdom and elsewhere.

This question was due to be discussed at the Debate Tribune at the Institute of Physics, Belgrade, on 25 June, but the debate was cancelled by management for fear that the truth about the situation would come out. The truth is that many research institutes are functioning at less than half capacity because employees are running private businesses, or are engaged in other nonscientific activities, while retaining their scientific posts even though they do not turn up for work regularly. Nobody has been removed from projects or fired for this corrupt practice, which is becoming widespread. This is why some people advocate placing fundamental research, which is hard to control administratively, in universities, where staff earn their pay by teaching even if they neglect their research activities.

But the principal problem is psychological. After 15 years of despair and demoralization, it is not easy to persuade people to return to their professional duties. In particular, many institutes are run by people who came to power during the Milosevic era, and it will probably be another generation before Serbia is decontaminated in this respect.

Petar Grujic

Institute of Physics, PO Box 57, 11080 Belgrade, Serbia, Yugoslavia

Shadows of Plato's genes

Regulatory gene pathways could explain animals' different morphologies.

The Evolution of Developmental Pathways

by Adam S. Wilkins

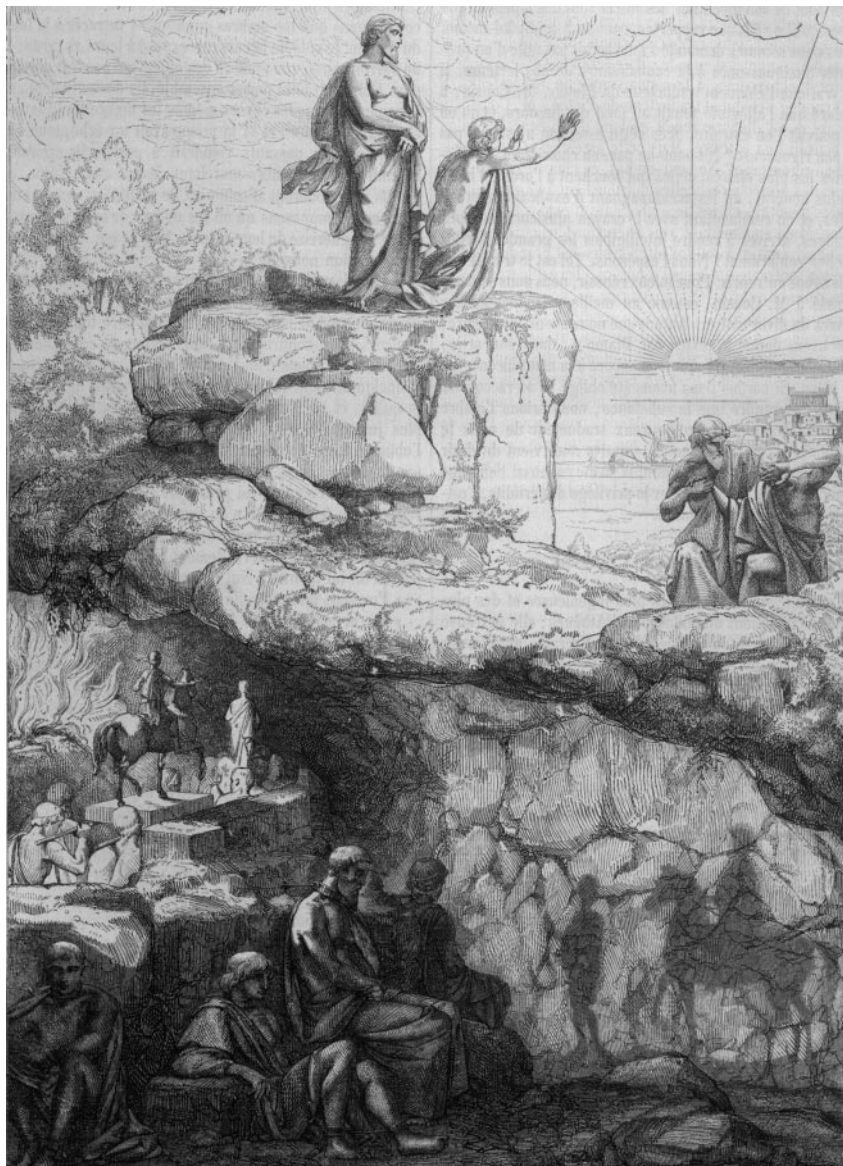
Sinauer Associates: 2002. 575 pp. \$54.95

Rudolf A. Raff

The rapid maturation of evolutionary developmental biology into a well-defined discipline is evident in the type of books being written about it. Until recently, authors sought to establish the boundaries of a new field, using a wide range of examples to illustrate phenomena and approaches to mechanisms linking two disparate subjects. But as the field matures, authors are beginning to assume the basic existence of the larger discipline and to focus on the subdisciplines. Adam Wilkins, editor of the review journal *BioEssays*, has now written a fascinating and lucid book that represents a transitional species between these approaches. The book builds a reductionist structure that points to where much of the discipline will be moving in its thinking and work: the evolutionary extension of developmental genetics.

Plato's *Republic* contains an allegory of people chained in a cave, who see only the flickering shadows of reality on the wall, until some are able to break away and see the underlying truth. This is similar to Wilkins' message on the role of developmental pathways in the evolution of development. Wilkins takes on the paradox that morphologically disparate creatures use conserved regulatory genes, and asks how different developmental processes can have evolved using much the same molecular machinery. He argues that the regulatory gene pathways are the reality under the morphology that we see and want to explain. His perspective is powerful and compelling. Although it may be too early to decide what constitutes the fundamental reality in the interactions between evolution and development, the book provides the framework for a coherent research programme that should be highly effective.

The book contains some familiar topics to set the evo-devo perspective. These include a historical summary, a bit about the fossil record and a discussion of phylogeny. It then progresses rapidly to the core issues by using a few concrete examples of changes in gene function in evolution, introduces the reader to developmental pathways and their functions, and then moves on to conserved genes and their evolutionary roles. These are all strongly presented themes that are essential to the ultimate argument, and Wilkins effectively introduces concepts of networks and the co-option of genes for new functions.



In Plato's *Republic*, people could see only the shadows of reality on the wall, not the underlying truth.

Much of the core of the book comprises detailed discussions of four major developmental pathways: sex determination, segment patterning in insects, vulval cell patterning in nematodes, and vertebrate limb development. I quailed when I contemplated reading these chapters, dreading the boring prose that detailed descriptions of such pathways often entail. Not here. Wilkins' writing is vibrant, and he is not afraid to address difficult questions with interesting ideas. Understanding the evolution of what seem to be well-integrated genetic pathways is difficult. But Wilkins is highly effective in discussing genetic

sources for evolving developmental systems, including transcription control sites, gene duplication and functional divergence, and gene recruitment. Finally, he argues forcefully that the optimization of pathways under selection occurs by the co-option of genes for new compensatory functions.

Vertebrate limb development, which is the least well understood of the pathways that Wilkins discusses, provides a striking intersection of evolutionary patterning and developmental pathways. The tale is nicely told, with an intriguing exploration of how changes in genetic architecture, for example of *Hoxa* and *Hoxd* paralogues, are translated

into observed evolutionary changes in limb features, such as digit number. The story includes some interesting considerations from the morphological evolution of limb components, but to get the complementary view from a palaeontological perspective, readers also should take a look at Jennifer Clack's new book, *Gaining Ground* (Indiana University Press).

In the final chapters, Wilkins discusses some interesting but poorly understood problems that continue to pose conundrums. These include morphogenesis, metazoan origins and the relationship between microevolution and macroevolution. The latter question is about whether the processes seen on the microevolutionary scale are commensurate with the large effects seen on the macroevolutionary scale. The gap is narrowing as the roles of mutations that have large effects are becoming better understood, and this synthesis will be as important as any in evo-devo.

The book has relatively few weaknesses. In general, it has too few citations, so some major contributors are not mentioned. Wilkins unfortunately gives some comfort to the concept of 'functional homology', which is an unhelpful confounding of the definition of homology with criteria for its recognition. And the discussion of the discrete phases of Hox expression, so crucial in tetrapod limb development, cries out for a figure. But these deficiencies are of minor consequence — this is a splendid book. Wilkins has produced a lively, readable and thoughtful work that will help readers to see the discipline with new rigour and provide a source for stimulating discussions in evo-devo courses. ■

Rudolf A. Raff is in the Department of Biology, Indiana University, Bloomington, Indiana 47405-7107, USA.

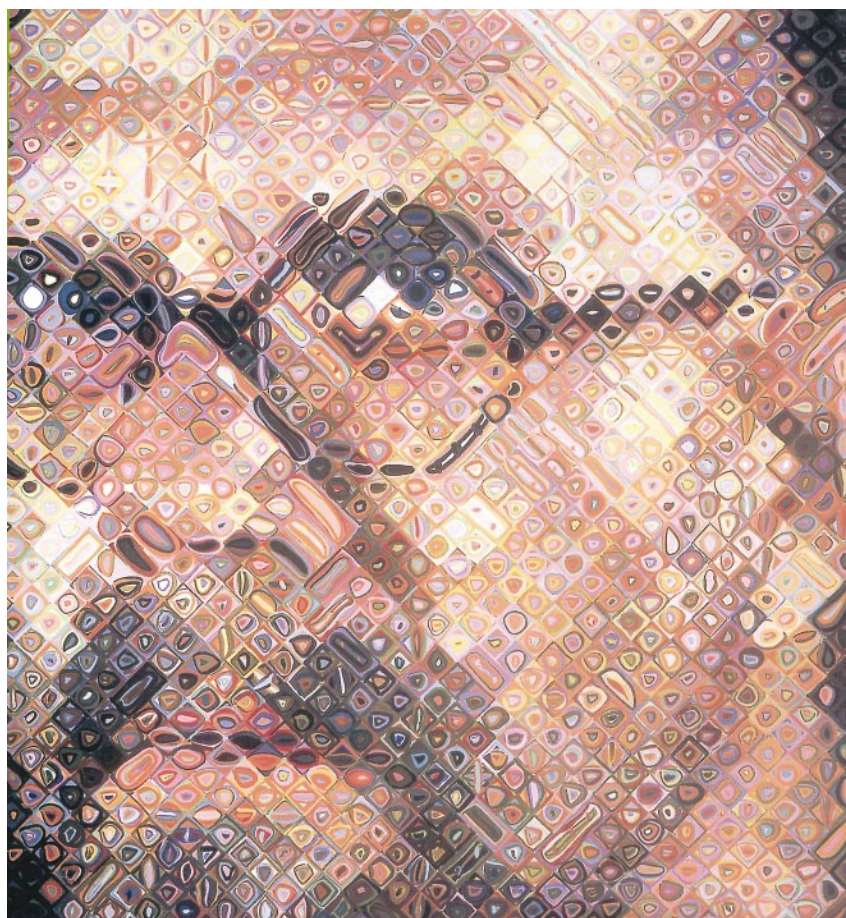
Trying to make sense of art

Vision and Art: The Biology of Seeing

by Margaret Livingstone
Harry N. Abrams: 2002. 208 pp. \$45

Semir Zeki

Artists differ hugely in their styles and creations, and their art appeals to some and not to others — which is another way of saying that one of the characteristics of art is its richness and variety. This no doubt reflects the variability in the organ that creates and appreciates art, the brain. Yet no one has been able to relate the variability in artistic creativity and appreciation to any given brain structure or process, partly because no one knows the neural processes underlying



Seeing the big picture: blocks of different colours are viewed together in Chuck Close's *Self-Portrait*.

the creative impulse or brain variability. And yet these differences in brain organization, whatever they may turn out to be, are superimposed on a common plan that is characteristic of all brains. It is this common organization that allows us to communicate through art and about art without using the written or spoken word.

This simple message, not by any means a new one, is worth repeating. Its repetition in this extensively illustrated book by Margaret Livingstone, and in the foreword by David Hubel, is to be welcomed. The book may provide yet another stimulus for neurobiologists of vision to study more closely what artists produce, and to use art in trying to further our understanding of the principles underlying the organization of the visual brain. Visual artists are, in a sense, neurobiologists of vision, studying the potential and capacity of the visual brain with techniques that are unique to them. This book illustrates that point well.

Livingstone uses one of the most conspicuous features of the organization of the visual brain — the functional specialization of different areas within it for processing different attributes of the visual scene, such as motion, colour, form and faces — to explain how we can account perceptually for much

that we see in paintings. Not surprisingly, there is more in the book about the colour, motion and depth systems than about the form system, because they are better understood in neurobiological terms. The many illustrations clarify some of the points very well. For example, page 34 provides a convincing answer to the often-asked question of how the colour-blindness that results from cerebral lesions differs from inherited retinal colour-blindness.

Although the text should be accessible to many readers, the parts that relate to the cortex are rather simplistic and ultimately misleading. Many today would not agree with the effort to partition the whole of the visual apparatus, of visual perception and of visual art, into two subdivisions of the brain, the 'what' and the 'where' systems. It is hard, for example, to accept the supposition that cubism is extreme in its spatial imprecision because it is exclusively the product of the spatially imprecise 'what' system. Cubism constitutes a deliberate effort to de-emphasize unique points of view and unique viewing conditions. It almost certainly taps deeply into many of the specialized systems, including the frontal lobes.

It is also unfortunate that not all of the illustrations chosen to make physiological

points are perceptually convincing. This makes the book less compelling than it would otherwise have been. I do not see the poppies in the Monet painting reproduced on page 152 as moving or changing position, nor can I detect an illusory motion in the ice floes on page 161. There are several such examples in the book of a conflict between what the text says and what the visual brain experiences.

The coverage itself also sometimes causes problems. It is not easy, for example, to accept the proposition that one's gaze is drawn to the face or eyes in a painting because these are painted in greater detail than the surroundings. There are many counter-examples, unfortunately not given here, where the face is very sketchily drawn but still constitutes the immediate focus of attention. The reason for this is simple. The face carries more information and, correspondingly, the brain has devoted a specialized area to the processing and perception of faces.

There are, happily, many other illustrations that are perceptually much more convincing. Prominent among these are the many visual illusions, some well known and others less so, that are scattered throughout the book. Collectively, these should give the visual scientist a great deal of interesting material to think about. Above all, I hope that through works such as these, visual scientists will come to realize what a rich resource they are provided with by artists who exploit the potential of the visual brain in their creations. Such material is well worth studying scientifically. ■

Semir Zeki is in the Wellcome Laboratory of Neurobiology, University College London, Gower Street, London WC1E 6BT, UK.

A rounded view of Faraday

Faraday: The Life

by James Hamilton

HarperCollins: 2002. 448 pp. £25

Jacqueline Reynolds & Charles Tanford

In the introduction to his monumental work on the life of Isaac Newton, *Never at Rest* (Cambridge University Press, 1981), Richard Westfall asserted his desire to avoid a detailed factual essay on newtonian science. He chose instead to make Newton himself, the scientist and the man, the central character in the drama of his unfolding life. Treading this fine line between yet another account of well-known scientific discoveries and providing a portrait of the individual who made such discoveries is a difficult path, which Westfall negotiated with great skill.

Hamilton's life of Michael Faraday is of

this genre. It is a portrayal of Faraday from his early days as an enthusiastic, observant young man through to his declining years — a man living actively through the intellectual and industrial explosion of the Victorian age. His friends and associates included not only other scientists, but many of the artists and literati of the day. All the influences of the time and the intellectual developments in Britain and mainland Europe were part and parcel of Faraday's life. The 'drama', as Westfall would have it, unfolds through the medium of his correspondence, laboratory notebooks and writings that were intended only for his own use (which today we might categorize as diaries, although they were far more extensive than that commonplace word suggests).

The strength of this book lies in the carefully constructed characters surrounding Faraday: the not always admirable Humphry Davy and his appalling wife; the young men who were part of Faraday's youthful ventures into science; his supportive but non-intellectual wife; unusual women such as Mary Somerville and Ada Lovelace, with whom Faraday had intellectual and emotional encounters. Faraday himself becomes a more real and sympathetic individual in this book than in some other accounts that concentrate primarily on his science or his unusual religious background as a member of the fundamentalist Sandemanian sect.

Hamilton reiterates and provides support for Faraday's formal devotion to the strict rules of this sect, as described by other writers (notably Geoffrey Cantor), but here Faraday's day-by-day activities reveal another side of him. As a young man on his European tour with Davy, he is described as a "*bon viveur*, enjoying good food and wine"; 35 years later he still enjoyed regular visits to the theatre.

Less satisfying is the presentation of



Bon viveur: Michael Faraday had strict religious beliefs, but also enjoyed good food and wine.

Faraday's prodigious achievements in the areas of analytical chemistry, electricity and magnetism, optics and organic chemistry. To have matched the more detailed and insightful essays on this subject that exist elsewhere in the Faraday literature would have required a much longer book and perhaps greater familiarity with these specific scientific areas. Hamilton's descriptions of experiments and their outcomes are marvelously pictorial and often humorous, relying heavily on Faraday's own notebooks — the reader might almost have been there through the course of numerous explosions and emanations of noxious fumes. But Hamilton, whose earlier work includes a biography of the landscape painter J. M. W. Turner, is less sure-footed when placing Faraday's achievements in a historical perspective.

That said, there are plenty of 'scientific essays', but it is hard to find a sympathetic and intriguing account of the life of a great scientist. If people and their peculiarities interest you less than the science, this is probably not the book for you. But if you enjoy having the great and the good from the distant past brought back to life (warts and all), then this is recommended reading. ■

Jacqueline Reynolds and Charles Tanford are at Tarlswood, Back Lane, Easingwold, York YO61 3BG, UK.

The temperature of everything

A Matter of Degrees: What Temperature Reveals About the Past and Future of Our Species, Planet, and Universe

by Gino Segre

Viking: 2002. 320 pp. \$24.95

Vaclav Smil

Book publishing, like designer clothes and the most-cited scientific articles, follows fashion. Generic-topic books are now definitely in. The recipe is simple: pick an object or an idea and examine it from just about any conceivable angle. The result is a crop of books that includes volumes on dust and on longitude, on fire and on childhood, on salt and on excrement (yes, the title of the translation from the French original is a four-letter word). And, although not quite identical, the pitch of these books is clearly similar — they aspire to be science 'lite', accessible to the masses and hence fashionably promoted. *A Matter of Degrees* is one of the latest additions to this popular genre; the thread, in the author's words, is temperature.

As an incorrigible interdisciplinarian I am generally positively predisposed towards these amoebic topics. So why am I not

robustly enthusiastic about this book? Certainly not because of its quality — the author, a theoretical physicist and well-known contributor to high-energy research, skates on the solid ice of his own discipline, and has diligently studied the biological literature, which is essential to understanding the book's first and fourth chapters, and used it accurately. And also not because the book is a chore to read, for it is well written, interesting and informative. And, needless to say, the book's big topic is about as fundamental, and hence as important, as it gets in studies of both the living and the inanimate Universe.

This may actually be the book's main problem. Books about longitude or dust have a huge area of reality to explore, and the topics lend themselves to pursuing some fascinating connections and consequences. Yet, as wide-ranging as they are, their scope is inherently limited. Enzymatic catalysis is unaffected by longitude, and Saharan dust is irrelevant to the radioactive decay in the Earth's crust. But what does not fall under the influence of temperature? What object or process can be seen and studied as devoid of it or unaffected by its change? So a book about temperature really aspires to be a book about everything.

This is clearly impossible, but Segre

cannot be faulted for not trying. He crams in a great deal yet, inevitably, even some fundamental realities get no mention at all: temperature's distribution on the Earth and its effect on the evolution of life and on the fate of ecosystems are among the most obvious omissions. (An important aside: Segre uses almost exclusively annoying, antiquated units; of course, as this is a US book intended for a wide readership, the publisher may have insisted on Fahrenheit and feet, but I feel that the author should have resisted.)

And so the book opens with: "Ninety-eight point six. It is extraordinary how alike we are," and goes on to discuss thermoregulation in organisms and its problems. We are then offered chapters dealing with the history of temperature measurement, the origins of thermodynamics, temperature and the Earth (including the greenhouse effect) and life in extreme environments. Finally, there are brief excursions into astrophysics and studies of low-temperature and other superconductivity. Along the way the text encompasses mammalian scrotal sacs, the precession of the Earth's rotation axis, hydrothermal vents at the bottom of the Pacific Ocean, magnetic fields of rapidly rotating neutron stars, foraminifera (badly misspelled in the original)

and superconducting using YBCO (yttrium barium copper oxide).

People who make a fleeting appearance include, predictably, many icons of twentieth-century physics, including the author's famous uncle Emilio; Subrahmanyan Chandrasekhar, puzzling out the death of stars; Leo Szilard, designing a better refrigerator with Einstein; and Arno Penzias and Robert Wilson, measuring background radiation. Additions to this *mélange* of topics and people can be made *ad libitum*, as there is nothing that qualifies for exclusion.

This book is being promoted in the United States by national print and radio interviews to show how such an everyday concept as temperature "underlies our understanding of how the Universe works". As I have no illusions about the scientific literacy of the US public, I doubt that they will appreciate some of the finer points from the book. Only those with a fairly solid grounding in several key scientific disciplines will appreciate this book for its brisk and wide-ranging treatment of many fundamental temperature-related concepts. Those less well equipped will see too many trees in such an infinite forest.

Vaclav Smil is in the Department of Geography, University of Manitoba, Winnipeg, Manitoba, R3T 2N2, Canada.

Science in culture

Life in a drop of water

A display of sculptures by Sarah Parker-Eaton and Louise Hibbert was inspired by plankton.

Liz Hoggard

For the past 18 months, jeweller Sarah Parker-Eaton and woodworker Louise Hibbert have been working together on a series of hand-sized sculptures and boxes inspired by the microscopic world of plankton. The works are mostly fashioned from sycamore, with decorative details in silver, resin and acrylic ink.

The two artists have been collaborating with biological oceanographer David Thomas of the University of Wales, Bangor, who shares their enthusiasm for plankton. All three acknowledge their debt to the nineteenth-century German zoologist Ernst Haeckel — in particular for the outstanding images in his book *Kunstformen der Natur*. Haeckel was both an accomplished artist and scientist, whose somewhat stylized representations of

planktonic organisms informed part of the Art Nouveau movement. René Binet's design for the elaborate main entrance of the

1900 World Exposition in Paris, for example, made direct reference to Haeckel's drawings of radiolarians.

Thomas provided the two artists with the opportunity to explore under the microscope the shapes, forms and sublime movements of different types of living plankton, including diatoms and dinoflagellates, as well as radiolarians. Parker-Eaton says that this opened up for her "a miraculous underwater world. We never realized that a drop of water could contain such a variety of life."

Planktonic organisms have myriad forms. They can be multicoloured spheres or cubes, with sturdy spines or fine, elongated needles that deter predators. Some form delicate spiralling chains, whereas others, such as dinoflagellates, have whip-like flagella to help them move in the water. The artists try to reflect this richness in their work. In one sculpture, for example, they represent armour-plated dinoflagellates as tactile spheres made from ancient bog oak covered in heavy silver plates.

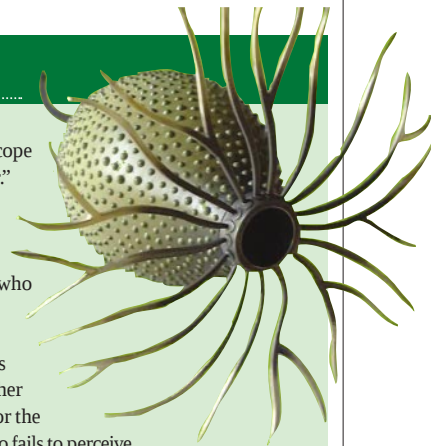
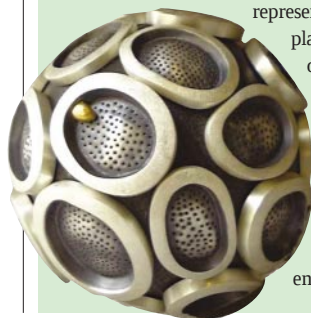
The vivid colour that Hibbert and Parker-Eaton use in their work highlights the extraordinary ability to refract light of planktonic organisms such as diatoms and radiolarians, whose cell walls are constructed from silicate. "The walls are basically glass," says Thomas. "And they are etched with intricate patterns, which under the microscope refract light, like crystal,

into a kaleidoscope of colour."

Thomas believes that a scientist who fails to see the aesthetics in his or her science, or the artist who fails to perceive (often without realizing it) the mathematical form in their subject, "will always fall short in their respective interpretations of nature's design".

Liz Hoggard is a freelance journalist based in London.

Plankton can be seen at the Crafts Council Shop at the V&A Museum in London until 15 September, and will be at Chelsea Crafts Fair at Chelsea Old Town Hall in London on 15–20 October.



A lethal side-effect

Linda Partridge and David Gems

Industrialized societies throughout the world are greying. Since 1840, maximum life expectancies have increased at a rate of about three months per year and this trend shows no sign of slowing down. The good news is that people are getting healthier. But one downside is the net impact on healthcare. The overall improvement in health is more than countered by the much greater number of individuals reaching ages at which age-related health problems occur. An obvious example is Alzheimer's disease, which was almost unknown a century ago. The same is true of age-related macular degeneration, now the leading cause of blindness. Ageing is bad for us and yet it happens to everyone. So why does it occur at all?

The effects of ageing are particularly obvious in humans, but are not peculiar to us. Ageing occurs in natural populations — as individuals get older they become less fecund and more likely to die. Organisms ranging from yeast to mammals to plants are affected. Cars and washing machines wear out too, which suggests that ageing could be an inevitable consequence of complexity. But at least some things do not age. All organisms living today are descended from lineages that have been going strong for three billion years. Germ lines do not wear out. So if ageing is not inevitable, surely such a universal and ultimately lethal process must have a purpose?

August Weismann suggested that ageing

functions to rid the species of worn out and decrepit individuals so as to reduce competition for resources with younger ones. The most obvious problem with this idea is that it is circular because it assumes the existence of the trait whose occurrence it is aiming to explain. The circle could be broken by viewing the inevitable accumulation of damage during a lifetime as an intrinsic trait that has evolved to increase the death rate of the elderly.

But even without the circularity, the claim that a trait has evolved for the benefit of a species requires close inspection. Most traits, such as eyes and digestive systems, evolved because they increase the survival and reproductive success of their bearer or, in a few cases such as mammary glands, the bearer's closest relatives. Fairly restrictive conditions are required for a trait to evolve because it benefits other members of the species. Explanations for the prevalence of ageing based on species-level functions such as increasing the rate of evolution have therefore fallen into disrepute.

Ageing is caused by the accumulation of damage, and no gene has evolved specifically to cause damage and debility. Yet genes do influence the rate of ageing. Birds live longer than comparably sized mammals, which suggests that the rate of ageing evolves. And mutations in single genes can increase the lifespans of laboratory animals. Many of these genes have been identified and are known to encode normal constituents of cells and endocrine systems. So why has natural selection favoured the wild-type

Ageing

Ageing is bad for us and yet it happens to everyone. So why does it occur at all?

form of the gene rather than a mutant trait that extends lifespan?

Genetic effects on ageing can be understood only as a side-effect of something else. Genes that slow ageing could exert these effects because they repress the causes of ageing-related damage. Reproduction seems to be one of these sources of damage, because fecundity is often reduced both during the evolution of slow ageing and by single-gene mutations that extend lifespan. Food seems to be another damage source, because many of the genes that slow ageing are involved in the response to changing nutrient levels. And reducing food intake slows down ageing in organisms ranging from yeast to mammals.

So adaptive, programmed events that are controlled by genes cause damage and therefore lead to ageing as a side-effect. Ageing is not a programmed process like development. No hierarchy of genetic control systems has evolved to ensure that ageing occurs in the right place and at the right time. It is a late-onset genetic disease that affects all of us, a result of damage inflicted by other, adaptive processes earlier in life.

The steady increase in life expectancy in human populations shows that longevity is plastic. Although lifespans are species-specific, they can be greatly modified by the environment as well as by genes. For many human populations, the fixed three score years and ten allotted for human longevity are already but a distant memory. Much of this increase in lifespan has been achieved by improvements in public health, medical care and domestic circumstances. We are beginning to view ageing-related damage as a side-effect of other, adaptive processes. This may allow us to reduce the impact of ageing-related diseases as the limits on human lifespan recede. ■

Linda Partridge and David Gems are in the Department of Biology, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK

FURTHER READING

Austad, S. N. *Why We Age* (Wiley & Sons, New York, 1997).
Ricklefs, R. E. & Finch, C. *Ageing: A Natural History* (Sci. Am. Library, New York, 1995).
Partridge, L. & Gems, D. *Nature Rev. Genet.* **3**, 165–175 (2002).

“One man in his time plays many parts”: the idea of *The Ages of Man* is a familiar one, but is ageing more properly viewed not as a process in its own right, but as a side-effect of accumulated genetic damage?

BRIDGEMAN



Inside the professionals

Jonathan W. Yewdell and David C. Tschärke

Real-time microscopy is providing fresh insights in many fields of biology. Immunology is no exception, as demonstrated by striking images of the inner workings of dendritic cells.

When microbes attack vertebrate organisms, dendritic cells help to raise the alarm in the immune system. Two papers in this issue, by Boes *et al.*¹ and Chow *et al.*² (pages 983 and 988), provide an unprecedented inside view of how these cells go about their business.

Dendritic cells are both sensors and messengers that form an early-warning network in many tissues of the body. On meeting a pathogen, they 'mature' and migrate to the nearest lymph node, carrying bits and pieces of the invaders. Inside dendritic cells, these foreign proteins (antigens) are diced into short peptides and loaded onto cell proteins, termed MHC class I or class II molecules, whose job is to take the fragments back to the cell surface and present them to T cells (Fig. 1). T cells that recognize these fragments of pathogen are then activated and join the battle against the invader, where they secrete a range of powerful molecules that summon and organize more immune cells, as well as directly harming the microbe.

Many cell types can present antigens to already activated T cells as an extra job on the side, but dendritic cells are needed for the original activation of T cells, and this is their only task. In immunospeak, dendritic cells are professional antigen-presenting cells (pAPCs), and there is considerable interest in exposing their secrets. Boes *et al.* and Chow *et al.* have used real-time microscopy to visualize the remarkable alterations in the trafficking of MHC class II molecules that occur following dendritic-cell maturation.

Both groups started by genetically fusing green fluorescent protein (GFP) to the tail of class II molecules. Happily, GFP is highly fluorescent in cells but has little effect on the proteins it tags: the manufacturers of this marker owe a considerable debt to the humble jellyfish in which it evolved. Boes *et al.* genetically engineered mice so that all of their MHC class II molecules had glowing tails, and showed that this modification had no discernible effect on class-II-dependent immune functions. Because pAPCs are the only cells that always express class II molecules at high levels, they could be visualized in living tissue without further manipulation. Using confocal microscopy, Boes *et al.* provide a dramatic three-dimensional picture of the distribution of dendritic cells in living skin — an image sure to grace future immunology textbooks.

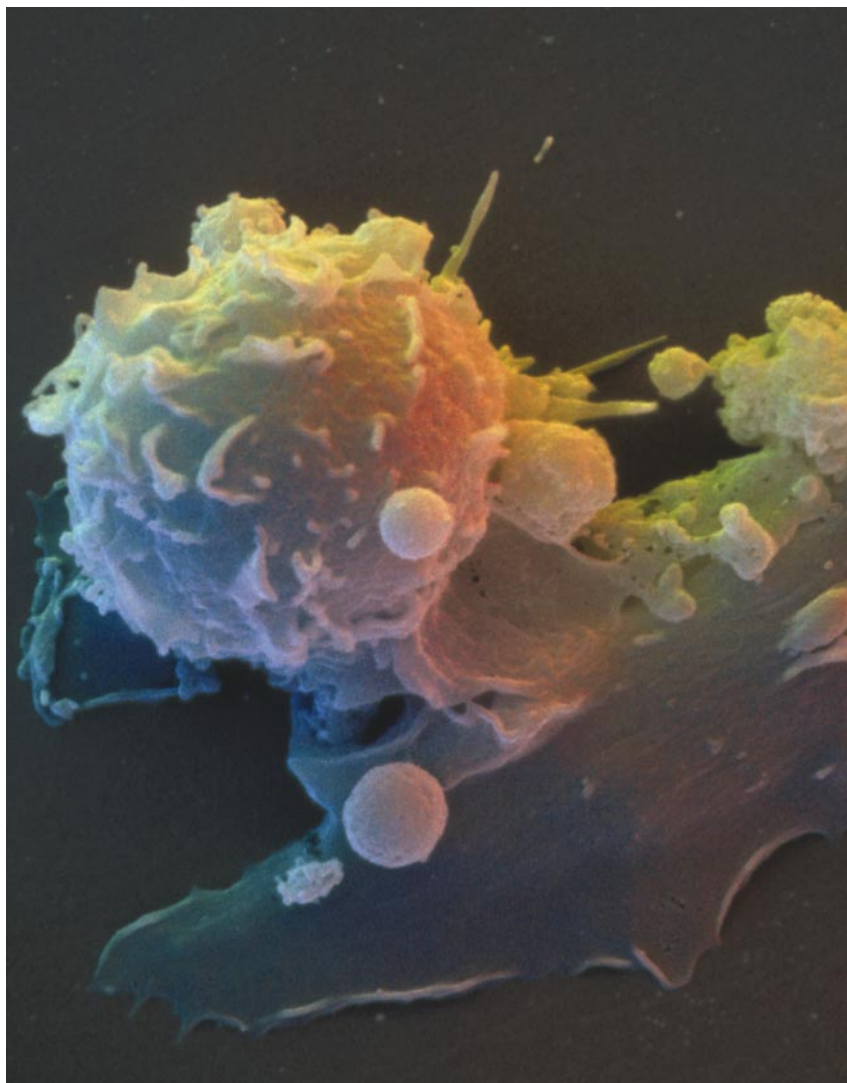


Figure 1 Immune interactions: a T cell (upper left) meets a dendritic cell.

Further experiments were even more illuminating. In these, Boes *et al.* visualized the intracellular movement of the GFP-tagged class II molecules by time-lapse imaging of dendritic cells that had been cultured from bone marrow and then activated by exposure to bacterial glycolipids. Immature dendritic cells energetically sample their external environments, but are lethargic in loading their class II molecules with the fragments of proteins they devour. The largely undigested meal is stored with class II molecules in the MHC class II

compartment (MIIC), a specialized form of an organelle called the lysosome. Upon activation, the dendritic cells stop eating, and concentrate on generating complexes between peptides and MHC class II molecules for T-cell activation.

Chow *et al.*² used a retrovirus vector to express GFP-tagged class II molecules in cultured dendritic cells. They show that, within 30 minutes of adding bacterial glycolipids, tubules emanate from the MIIC — which lies near the microtubule-organizing centre (the central station for a system of fibres used in

D. SCHÄRKE/SPL

part as an intracellular railroad) — and travel to the cell's outer plasma membrane. This observation extends the elegant immunoelectron microscopic findings of Kleijmeer *et al.*³ to living cells. Using total-internal-reflectance fluorescence microscopy, which allows visualization of events occurring at the cell surface, Chow *et al.* directly demonstrate that MIIC-derived vesicles deliver class II molecules to the plasma membrane. This is the first such demonstration of centrifugal movement of integral membrane proteins from lysosome-derived structures to the plasma membrane. Although this route is known only in dendritic cells so far, it is likely to exist in other cells.

Most remarkable is Boes and colleagues' demonstration that class II transport in dendritic cells can be influenced by T-cell recognition of antigen on their surface. Within minutes of coming into contact with T cells that recognize an antigen being presented, class II molecules are directed from the MIIC by extraordinarily long (up to 50 μm) tubules along the microtubule railroad, right to the interface between the dendritic cell and the T cell. When a dendritic cell interacts with several T cells at the same time, many tubules form simultaneously to deliver class II molecules to each T cell.

The interface between a T cell and an APC has been termed the immunological synapse^{4,5}, in recognition of the specialized molecular communication between the two cells. Synapse formation induces many changes in T cells, including polarization of the secretory apparatus towards the synapse and reorganization of numerous cell-surface proteins at the point of synapse formation. The new findings show that alterations occur on the APC side of the synapse as well — begging the question of what else might be going on in dendritic cells (and other pAPCs) while they embrace T cells.

The finding that only T cells that recognize their antigen induced tubule polarization raises a conundrum: if there is already enough MHC–antigen on the surface of the dendritic cell to allow recognition by the T cell, why transport more of it to the synapse? Perhaps the original signal is weak and needs to be bolstered by the delivery of more complexes. Or perhaps the GFP-tagged class II molecules are a 'green herring', and the important cargo transported to the synapse is something else that resides in the MIIC. More generally, it will be of interest to visualize the gyrations of pAPCs and their class II molecules in their natural surroundings — the skin, mucosae and lymphoid organs of living animals. Such 'vital microscopy'⁶, which is now possible through two-photon confocal microscopy, is bound to revolutionize our understanding of the immune system. ■

Jonathan W. Yewdell and David C. Tschärke are in the Laboratory of Viral Diseases, National

Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20892-0440, USA.

e-mails: jyewdell@nih.gov
dtscharke@nih.gov

1. Boes, M. *et al.* *Nature* **418**, 983–988 (2002).

2. Chow, A., Toomre, D., Garrett, W. & Mellman, I. *Nature* **418**, 988–994 (2002).

3. Kleijmeer, M. *et al.* *J. Cell Biol.* **155**, 53–63 (2001).

4. Krummel, M. F. & Davis, M. M. *Curr. Opin. Immunol.* **14**, 66–74 (2002).

5. Dustin, M. L., Bromley, S. K., Davis, M. M. & Zhu, C. *Annu. Rev. Cell Dev. Biol.* **17**, 133–157 (2001).

6. Von Andrian, U. H. *Science* **296**, 1815–1817 (2002).

Planetary science

Birth of a Solar System

A. G. W. Cameron

Radioisotope dating of meteorites suggests that planets formed in the Solar System over shorter timescales than had been thought. There are consequences for how the Moon formed, but is this the final word?

Fortunately for those who attempt to trace the origins of our Solar System, when the primitive solar nebula and planetary bodies began to form there were radioactive nuclei with a wide range of half-lives present. The detection of these nuclei and their decay products, particularly in meteoritic rocks, is a useful tool for working out when various features of the planetary system formed.

Two papers in this issue, by Yin *et al.*¹ (page 949) and Kleine *et al.*² (page 952), describe dating experiments on a selection of meteorites, based on the hafnium nucleus ¹⁸²Hf, which decays to tungsten, ¹⁸²W, with a half-life of about 9 million years. This decay is particularly useful for estimating the time at which the cores of planets formed, because undecayed hafnium was left behind in the planet's silicate mantle, but tungsten was absorbed into the planetary core. Both sets of authors have independently determined the initial ratio of ¹⁸²Hf to its stable isotope ¹⁸⁰Hf in the solar nebula, arriving at essentially identical values. But their ratio is less than half the previous value³, with important implications for our understanding of how quickly the planets grew and how soon their cores were established.

Yin *et al.* and Kleine *et al.* calculate core-formation ages based on isotope data from a range of meteorites, including the eucrite meteorites that are believed to originate from the large asteroid Vesta. Their results indicate that the cores of Vesta, Mars and Earth all formed sooner after the formation of the solar nebula than previous estimates had allowed. According to the new data, Vesta's core formed only 3 million to 4 million years after the birth of the Solar System, in comparison with the previous value of about 16 million years; the cores of Earth and Mars formed at about 29 million and 13 million years, respectively. But the situation for the Earth and Mars is complicated because their growth involves absorbing bodies similar to Vesta and then reprocessing the cores within them. Yin *et al.* suggest that a more

realistic picture is a continual core-formation process in Earth up to about the 29-million-year mark, or more.

The new estimates also put the formation of the Moon further back in time — both papers estimate its birth at between 25 million and 30 million years after the formation of the Solar System. The favoured explanation for the origin of the Moon is a giant impact between the growing Earth and a planetary body at least as massive as Mars^{4,5}. One model assumes that the Earth was half formed and had a double collision with a body twice the mass of Mars⁴; another model assumes that when the Earth was 90% formed, it was struck once, by a body similar in mass to Mars⁵. According to Wetherill⁶, the growth time for Earth was about 100 million years, so these new timing measurements favour the double-collision model, but many more of these giant-impact simulations are needed. In any case, the iron core of the impacting body crashes through the planetary mantle and directly joins the core of the Earth.

Both papers suggest that the planets accumulated mass faster than had been thought. But how rapidly should we expect large asteroids and planets to form in the solar nebula? Astronomical studies of star-formation processes indicate that new stars are formed when a localized, dense region (a 'core') in an interstellar molecular cloud collapses. The timescale for the collapse of such a core in its own gravitational field is more than a million years. But if there is a supernova nearby, the impact of its shock wave on the cloud core can shorten this timescale to a few tens of thousands of years^{7,8}. Furthermore, a supernova is the only astronomical object that can produce all of the radioactive isotopes known to have existed in the solar nebula but that are now extinct⁴.

Once the solar nebula was established, the interstellar grains that it contained collided and stuck together (the kinetic energy of collision is likely to be radiated away before a rebound can occur⁹). In about

1,000 years, bodies of size 10–100 m had grown⁹. Wetherill's simulations⁶ of subsequent planetary growth show that objects with the mass of Mars would have grown in about 100,000 years.

To form a core, the interior of the growing planet must be heated enough to melt the iron contained in its rocky mantle. This heat is generated partly through gravitational energy release and partly through radioactivity in the interior, particularly that of ²⁶Al (with a half-life of 740,000 years), which is expected to be the most important heat source for the smaller bodies such as Vesta. Accumulative collisions between these small bodies produce a planet. When they collide, the smaller body is splashed into the mantle of the larger body, and its molten-iron core (if formed) percolates down to the core of the larger body. So core formation is an extended process, and the formation time derived from samples of the rock is only an average. The two papers published here use three different models, so their differing core-formation times reflect the different parameters of those models.

However, another development in the study of now-extinct radioactive elements in the Solar System has an important bearing on these interpretations. A linchpin in these analyses is the assumption that millimetre-sized particles of calcium and aluminium silicates — calcium–aluminium inclusions (CAIs) — found inside some meteorites are the oldest objects in the Solar System, and that their formation marks the birth of the solar nebula.

But a recent discovery¹⁰ suggests that the CAIs might, in fact, predate the solar nebula: the one-time presence of the very short-lived nucleus ⁷Be (with a half-life of just 53 days) has been detected inside CAIs through abundance variations in its decay product, ⁷Li. This finding suggests instead that CAIs might have formed in an expanding supernova envelope and were injected, along with the radioactive isotopes, into the molecular-cloud core whose accelerated collapse formed the nebula (my own unpublished data; a preprint is available on request).

One consequence of supernova-produced CAIs is that the initial, solar-nebula abundance of radioactive ²⁶Al would have been much lower (because of dilution by the stable isotope ²⁷Al during the injection process). So it would have been a considerably weaker heat source in young planets than expected. Reducing the initial Solar-System abundance of ²⁶Al also affects the timescales of core formation: although Yin *et al.*¹ and Kleine *et al.*² predict an earlier core formation and faster accretion than had been thought, the formation times would be reduced still further. The core-formation time for Vesta might in fact be only 20–40% of the estimates published here. ■

A. G. W. Cameron is at the Lunar and Planetary

Laboratory, University of Arizona, Tucson, Arizona 85721, USA.

e-mail: acameron@lpl.arizona.edu

1. Yin, Q. *et al.* *Nature* **418**, 949–952 (2002).
2. Kleine, T., Münker, C., Mezger, K. & Palmer, H. *Nature* **418**, 952–955 (2002).
3. Lee, D.-C. & Halliday, A. N. *Chem. Geol.* **169**, 35–43 (2000).
4. Cameron, A. G. W. *Meteoritics Planet. Sci.* **36**, 9–22 (2001).

5. Canup, R. M. & Asphaug, E. *Nature* **412**, 708–712 (2001).
6. Wetherill, G. W. *Annu. Rev. Earth Planet. Sci.* **18**, 205–256 (1990).
7. Vanhala, H. A. T. & Cameron, A. G. W. *Astrophys. J.* **508**, 291–307 (1998).
8. Vanhala, H. A. T. & Boss, A. P. *Astrophys. J.* **538**, 911–921 (2000).
9. Cameron, A. G. W. *Meteoritics* **30**, 133–161 (1995).
10. Chaussidon, M., Robert, F. & McKeegan, K. D. *Lunar Planet. Sci.* **XXXIII**, abstract 1563 (2002); adswww.harvard.edu

Marine biology

Unveiling an ocean phantom

Vera L. Trainer

Certain episodes of mass fish mortality in coastal waters off the eastern United States have been ascribed to a planktonic organism called *Pfiesteria*. There are now fresh clues to how these fish are killed.

In James Powlik's novel *Sea Change*¹, a colony of mutant cells called *Pfiesteria* grows into a carpet-like monster stretching over several square miles of the Strait of Juan de Fuca, a body of water separating the west coast of the United States and Canada. Its slow movement south threatens the residents of Seattle with a nightmarish fate — death following inhalation or ingestion of a toxin produced by the microbes. The good news is that the *Sea Change* story is fiction; the bad news is that *Pfiesteria* is fact. But what is the truth about this 'phantom of the ocean'? It is thought that *Pfiesteria*, which belongs to a group of planktonic organisms called dinoflagellates, releases a toxin that ultimately kills fish². But using simple methods, Vogelbein and his colleagues (page 967 of this issue³) conclude that the toxin is not released into the environment and that the organism kills by direct methods.

In reality, the presence of *Pfiesteria* has not been documented along the US west coast but rather along the eastern seaboard, with a range from Delaware to Alabama⁴. It has been associated with massive fish kills,

especially off North Carolina². Because of the harm that results when *Pfiesteria* multiplies to high numbers, it is categorized with other plankton known to cause harmful 'blooms'. Many of these bloom-forming organisms produce potent toxins, so the assumption that *Pfiesteria* releases a toxin is not unfounded. Moreover, the population explosions of various single-celled plankton, including diatoms and dinoflagellates, are frequently associated with the production of nerve toxins — saxitoxin, brevetoxin, maitotoxin and the like⁵ — that are damaging when ingested by higher organisms.

These toxins are the key for researchers wishing to unlock the secrets of harmful algal blooms. By knowing the chemical identity of the toxins, highly specific detection methods can be developed to track the location of blooms and to discover where they originate, how they spread and when they are most toxic. The impact of harmful algae on living creatures, such as seabirds, marine mammals and even humans, can be assessed by determining the levels of toxin in their tissues and body fluids. But first the toxins

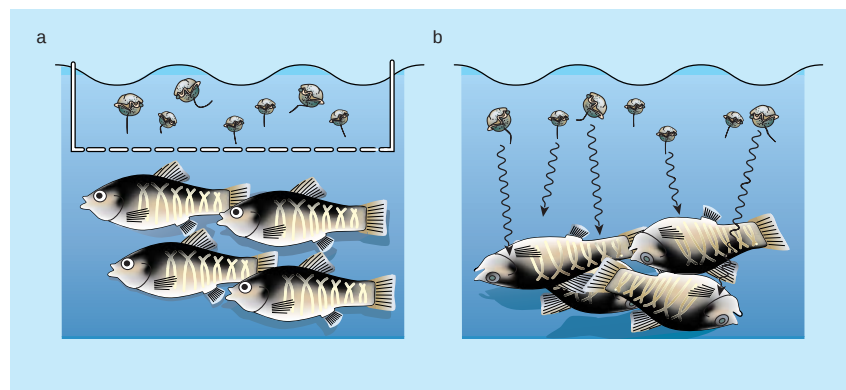


Figure 1 The dialysis technique used by Vogelbein *et al.*³. a, Where a membrane of pore size 3 μ m separates *Pfiesteria* cells from fish, no fish deaths are observed. b, However, fish die when they are in direct contact with the microbe. Separate control experiments showed that two typical algal toxins — saxitoxin (which is hydrophilic) and brevetoxin (hydrophobic) — can pass through the pores of the membrane. From these and other experiments, the authors conclude that *Pfiesteria*-induced fish kills are not due to the release of a toxin.

must be isolated from the clusters of cells that release them into the environment.

This is typically done by picking a single suspect cell from the mixture of microbes in a water sample from an estuary or ocean during a harmful event, and allowing it to multiply in the laboratory. Chemical constituents are then extracted from several litres of the cells, and toxic components are identified using a bioassay, a test that assesses whether a particular fraction is poisonous, usually to a fish or mouse. For almost a decade, researchers have tried unsuccessfully to identify, from many litres of isolated *Pfiesteria* cells, a toxin that could explain the mass mortalities observed in the wild⁶.

Vogelbein *et al.*³ and — in a related study, with some authors in common — Berry *et al.*⁷ have taken a different approach to the quest for a toxin released by *Pfiesteria*. First, they grew *Pfiesteria* in the lab and determined that the cultures could kill fish through direct contact. Then they separated the cells from the rest of the material using several methods, including filtration, centrifugation and dialysis (Fig. 1). They theorized that, if *Pfiesteria* cells release a toxin into sea water, this should be present in the cell-free fractions. But the experiments showed that the cell-free fractions do not cause fish death, and that *Pfiesteria* is lethal only when cells are in direct contact with the fish. The researchers therefore concluded that a toxin is not released into the surrounding sea water.

If they are not killed by a toxin, how do fish exposed to *Pfiesteria* die? It has been known for many years that *Pfiesteria* cells extend a suction-cup-like appendage called a peduncle to digest fish tissue². Vogelbein *et al.* go a step further and propose that fish die because *Pfiesteria* literally sucks the life out of them. It attaches to fish skin using the peduncle, extending finger-like protrusions called filopodia, then ingests cell matter from the fish. This parasitic feeding behaviour by *Pfiesteria* is detailed in high-magnification microscope images in Vogelbein *et al.*³, and in a video clip available in their Supplementary Information.

The new work promises to bring us closer to unveiling the true nature of this phantom of the ocean. Many loose ends remain, however, and it is possible nonetheless that a toxin is involved. Humans have suffered from memory impairment thought to stem from exposure to *Pfiesteria*⁸. Are these memory problems caused by a toxin, in aerosol form, produced by another organism that often coexists with *Pfiesteria*, as Berry *et al.*⁷ suggest? Do the rod-shaped granules that Vogelbein and colleagues identified in *Pfiesteria* (see Fig. 4 on page 969) contain toxins that are released only after the peduncle becomes attached? Do only certain *Pfiesteria* isolates produce toxins, as Burkholder *et al.*⁹ have proposed, indicating

that the *Pfiesteria* studied by Vogelbein, Berry and their colleagues were simply non-toxic strains? Can contamination by a fungus or other microbe explain why *Pfiesteria* cultures routinely grown in the presence of fish do indeed appear to produce a toxin¹⁰?

Assembling these pieces of the complex puzzle posed by *Pfiesteria* will require exceptional cooperation among researchers of differing expertise. The opportunity to discuss the outstanding questions and to establish new collaborations will engage scientists at the Tenth International Conference on Harmful Algal Blooms, to be held in St Petersburg, Florida, this October. ■

Vera L. Trainer is at NOAA's Northwest Fisheries Science Center, 2725 Montlake Boulevard East, Seattle, Washington 98112, USA.

Apoptosis

Sculpture of a fly's head

Jun R. Huh and Bruce A. Hay

Hox proteins are needed during development to produce body segments with different shapes and functions. In fruitflies, one Hox protein sculpts certain segments of the head by activating a cell-death-inducing gene.

Cell death occurs throughout the development of every animal, enabling excess cells to be eliminated, tissues to be sculpted, and cells or tissues that have outlived their usefulness to be removed¹. Much of this death occurs by an active 'suicide' process known as apoptosis. Quite a lot is known about the molecular signals and machinery that bring about apoptosis², but much less is understood about how the regulators of development access this machinery to determine where and when cells should die. Writing last week in *Cell*, Lohmann and colleagues³ revealed one way in which this happens.

In developing fruitflies (*Drosophila melanogaster*), almost all normal cell death requires some combination of three death activators, namely the proteins encoded by the *reaper* (*rpr*), *head involution defective* (*hid*) and *grim* genes⁴. These proteins, and their mammalian counterparts, promote apoptosis at least in part by de-inhibiting the cellular executioners — protein-cleaving enzymes known as caspases⁵. Of the fruitfly cell-death activators, *rpr* is particularly fascinating because its expression (transcription) is increased in every cell that is committed to die. So the control regions of the *rpr* gene constitute a site at which many different developmental and environmental cell-death signals are integrated (Fig. 1), and analysis of those regions should provide insight into how the signals promote death. This approach has, for example, been successful in helping us understand the

e-mail: Vera.L.Trainer@noaa.gov

1. Powlik, J. *Sea Change* (Random House, New York, 1999).
2. Burkholder, J. M., Noga, E. J., Hobbs, C. H. & Glasgow, H. B. Jr *Nature* **358**, 407–410 (1992).
3. Vogelbein, W. K. *et al.* *Nature* **418**, 967–970 (2002); advance online publication, 7 August 2002 (doi:10.1038/nature01008).
4. Burkholder, J. M., Glasgow, H. B. & Hobbs, C. W. *Mar. Ecol. Prog. Ser.* **124**, 43–61 (1995).
5. Baden, D. G. & Trainer, V. L. in *Algal Toxins in Seafood and Drinking Water* (ed. Falconer, I. R.) 49–74 (Academic, San Diego, 1993).
6. Moeller, P. D. R. *et al.* *Environ. Health Perspect.* **109**, 739–743 (2001).
7. Berry, J. P. *et al.* *Proc. Natl Acad. Sci. USA* **99**, 10970–10975 (2002); advance online publication, 5 August 2002 (doi:10.1073/pnas.172221699).
8. Grattan, L. M., Oldach, D. & Morris, G. *BioScience* **51**, 853–857 (2001).
9. Burkholder, J. M. *et al.* *Environ. Health Perspect.* **109**, 667–679 (2001).
10. Gordon, A. S., Dyer, B. J., Seaborn, D. & Marshall, H. G. *Harmful Algae* **1**, 85–94 (2002).

cell death that occurs during insect metamorphosis, which is dependent on the hormone ecdysone⁶, and cell death induced by DNA damage⁷.

One important class of developmental regulators consists of the Hox genes. Found in a wide range of species, these genes encode transcription factors that have major roles in creating specific segment identities along the head-to-tail (anterior–posterior) body axis during development, rendering each segment unique in its shape and function⁸. In fruitflies, for example, Hox genes determine which segments will produce legs, wings or antennae, and so on. They do this by activating and repressing the activity of particular target genes in specific spatial and temporal domains. Their importance is illustrated by what happens when they are mutated: one segment becomes transformed into another, resulting in one segment-specific body structure being replaced by another (a leg for an antenna, for example).

The central question about Hox gene function is how a single transcription factor can generate the many characteristics of a particular segment, which are the result of a variety of developmental processes such as cell division, apoptosis and tissue invagination. It has long been accepted that Hox genes must regulate the activity of many different target genes. But the identification of these targets, and an understanding of how changes in their activity lead to specific developmental outcomes, has been slow in coming.

Lohmann *et al.*³ have now tied the fields

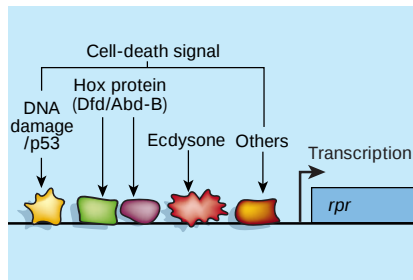


Figure 1 Focal point for the control of cell death: the reaper (*rpr*) gene is activated in all fruitfly cells fated to die. Activation requires particular gene-transcription factors (coloured shapes) to bind to specific sequences upstream of the *rpr* promoter (a control region). These factors and their associated binding sites have been identified for several different cell-death activators, including the insect steroid hormone ecdysone⁶, which triggers the destruction of tissues during metamorphosis, and DNA damage⁷, removing cells with potentially damaged genomes. Lohmann *et al.*³ have found that the Hox protein Dfd, and perhaps Abd-B, uses transcriptional activation of *rpr* to sculpt segment boundaries during development. It is likely that many other regulators of *rpr* transcription remain to be identified.

of development and apoptosis together, giving an insight into how Hox genes harness cell death to sculpt tissues in fruitflies. They show that the protein encoded by the Hox gene *Deformed* (*Dfd*) directly activates the expression of *rpr*, and that this is both necessary and sufficient to account for *Dfd*'s ability to maintain a segment boundary in the fruitfly head.

Lohmann *et al.* took a beguilingly straightforward approach to finding out which genes are targets for *Dfd*. They sought out gene mutations that result in embryonic defects similar to those caused by *Dfd* mutations, the idea being that the mutated genes might encode important *Dfd* targets. *Dfd* mutants have several prominent defects in the jaw region: they have too many cells in the ventral maxillary segment, and they lack the boundary between the maxillary and mandibular segments. Similar defects had been seen previously⁹ in fruitfly embryos lacking a chunk of DNA sequence in a chromosomal interval designated Df(3)H99. This interval contains the *rpr*, *hid* and *grim* genes, and embryos lacking the interval entirely show almost no normal cell deaths¹⁰. These findings, together with earlier observations that both cell death and *rpr* expression are prominent in the head^{9,10}, pointed to the idea that *Dfd* might carve out the boundary between the maxillary and mandibular segments by promoting *rpr*-dependent cell death.

Lohmann and colleagues have several results that support this hypothesis. First

they find that embryos with mutant *Dfd* show decreased apoptosis in the head, consistent with the idea that *Dfd* normally functions to promote cell death, whether directly or indirectly. Second, forced expression of DIAP1 — an inhibitor of *rpr*-, *hid*- and *grim*-dependent cell death — in the head region of wild-type embryos gives rise to defects similar to those seen in *Dfd* mutants. This observation is critical because it shows that reduced apoptosis alone can account for the head defects seen in the *Dfd* and Df(3)H99 mutant embryos.

Is the *rpr* gene a direct target of the *Dfd* protein? The authors argue that it is. For instance, the expression of *rpr* at the boundary between maxillary and mandibular segments requires the presence of *Dfd*, but the expression of *hid* and *grim* in the maxillary segment does not. Moreover, forced expression of *Dfd* can induce *rpr* expression and cell death. Finally, DNA sequences upstream of the *rpr* gene contain sites that bind *Dfd*; DNA fragments containing these sites are enough to drive the expression of a test gene at the maxillary-mandibular boundary in wild-type embryos. This requires a functional *Dfd* protein, as well as intact *Dfd*-binding sites.

So *rpr* is a direct transcriptional target of *Dfd*. But is the expression of *rpr* actually required to sculpt the fruitfly head? Lohmann *et al.* show that in one sense it is, because if *rpr* alone — and not *hid* and *grim* — is deleted, segment-boundary defects characteristic of *Dfd* mutants are seen during mid-embryo development. But in another sense (that of ultimate outcomes) *rpr* is not required, because *rpr* mutants hatch with normal head structures. The likely explanation is that *rpr* normally works together with other apoptosis inducers such as *hid*, which is also expressed in maxillary cells at several stages and is required for head development. So the loss of *rpr* delays, but does not eliminate, the crucial cell deaths.

Could this Hox-gene-dependent induction of cell death be a more general phenomenon? Several observations suggest that it might be. Lohmann *et al.* found that mutants showing abnormal expression of a second Hox gene, *Abd-B*, have a partial fusion of the abdominal segments in which *Abd-B* would normally be expressed. These mutants also show a decrease in *rpr* expression in cells at the posterior segment border, which would usually express *Abd-B*. So *Abd-B*-dependent expression of *rpr* could be important for generating or maintaining these segment boundaries, too. Moreover, several mouse Hox gene mutants have been described in which segment boundaries are not maintained¹¹, or in which normal apoptotic tissue sculpting fails to occur¹². Direct links between these Hox genes and components of the apoptotic machinery have not been documented. But the results of Lohmann *et al.*



100 YEARS AGO

While working on the reflective power of cyanin mirrors I have noticed some very interesting effects of light on that substance. Freshly fused cyanin is of a deep metallic bronze colour, but exposure to light turns it plum colour and finally a steely blue-black. In the moderate light of a cloudy day the change is perceptible in half an hour, in direct sunlight in less than a minute... By an exposure of thirty hours I have obtained on cyanin easily recognisable photographs of small, well-illuminated objects. A cyanin mirror, or better yet a piece of ground glass washed over with fused cyanin, exposed for ten hours to the spectrum of a Nernst lamp shows the effect to be very strong in the yellow, just perceptible in the adjacent red and green, and imperceptible in the blue and ultra-violet... At the same time, the exposure to light greatly decreases the absorbing power where it was originally large, as may be easily seen on looking at a sodium flame or a spectrum through an exposed coating of cyanin. It is as though the absorption were due to molecular resonance and the light produced a fatigue or destruction of this resonating power. From *Nature* 28 August 1902.

50 YEARS AGO

A brief survey of housing development in the U.S.S.R. from 1917 until the present day... shows that the material difficulties with which the Government has had to deal have been formidable. The housing resources taken over by the Soviet authorities after the Revolution were both quantitatively inadequate and in appallingly poor condition. Very little new construction took place until the late 1920's... so that the pre-revolutionary level of housing was restored less quickly than that of the other branches of the economy... In the present stringent housing situation, the Government has controlled the distribution of available housing resources, taking into consideration the priority claims of ex-servicemen, the dependants of war casualties and invalids, the partisans and their families. The relationship between work (type, quantity and quality) and the allocation of limited housing resources, already recognized before the War, has now been further strengthened... Housing is thus regarded as one of the instruments for increasing the productivity of labour and building up permanent staffs in industry. From *Nature* 30 August 1952.

suggest that it would be well worth looking for such connections.

*Jun R. Huh and Bruce A. Hay are in the Division of Biology, MC156-29, California Institute of Technology, 1200 East California Boulevard, Pasadena, California 91125, USA.
e-mail: haybruce@its.caltech.edu*

1. Jacobson, M. D., Weil, M. & Raff, M. C. *Cell* **88**, 347–354 (1997).
2. Hengartner, M. O. *Nature* **407**, 770–776 (2000).
3. Lohmann, I., McGinnis, N., Bodmer, M. & McGinnis, W. *Cell* **110**, 457–466 (2002).
4. Bangs, P. & White, K. *Dev. Dyn.* **218**, 68–79 (2000).

5. Shi, Y. *Mol. Cell* **9**, 459–470 (2002).
6. Jiang, C., Lamblin, A. F., Steller, H. & Thummel, C. S. *Mol. Cell* **5**, 445–455 (2000).
7. Nordstrom, W. & Abrams, J. M. *Cell Death Differ.* **7**, 1035–1038 (2000).
8. Mann, R. S. & Morata, G. *Annu. Rev. Cell. Dev. Biol.* **16**, 243–271 (2000).
9. Nassif, C., Daniel, A., Lengyel, J. A. & Hartenstein, V. *Dev. Biol.* **197**, 170–186 (1998).
10. White, K. *et al. Science* **264**, 677–683 (1994).
11. Gavalas, A., Davenne, M., Lumsden, A., Chambon, P. & Rijli, F. M. *Development* **124**, 3693–3702 (1997).
12. Stadler, H. S., Higgins, K. M. & Capechi, M. R. *Development* **128**, 4177–4188 (2001).

Renewable fuels

Harnessing hydrogen

Esteban Chornet and Stefan Czernik

Biomass can produce clean fuels and could be a vital, renewable energy source for the future. The demonstration of hydrogen production from biomass-derived molecules marks progress towards this goal.

Fossil-fuel stocks are a limited resource and, as the world's governments struggle to agree on a strategy to combat pollution and greenhouse-gas emissions, the search for clean, renewable energy sources has never been more intense. On page 964 of this issue, Cortright *et al.*¹ provide experimental evidence that simple biomass-derived molecules, such as glucose and glycerol, can be treated to produce hydrogen with reasonable efficiency. The authors suggest that, with some additional effort, their technique could also be technologically and commercially viable.

Cortright *et al.* demonstrate that glucose (the sugar used as an energy source in both

plants and animals) and glycerol (derived from fats) can be reformed in the aqueous phase in the presence of a platinum-based catalyst to produce H₂. The conversion takes place at moderate temperatures, around 225–265 °C, and at pressures of 27–54 bar — conditions that prevent steam formation and ensure that the reaction sequence takes place in the aqueous phase.

The authors propose that the mechanism of hydrogen production involves the rupture and rearrangement of the biomolecules' C–C and C–O bonds on the platinum catalyst, leading to the formation of intermediates. These can then produce H₂ by reacting with the abundant water present —

glucose was used at a water–carbon molar ratio of 165, but the authors indicate that ratios as low as 15 are possible. Simple hydrocarbons and carbon dioxide are also formed. The amount of gaseous H₂ produced as a proportion of the reaction products ranges from 36–50% for glucose to 51–75% for glycerol; and carbon conversion to gaseous products is 50–84% for glucose and 83–99% for glycerol. A yield of up to 80 g of H₂ per kilogram of catalyst per hour is possible.

Cortright *et al.* claim that their approach represents a significant departure from traditional high-temperature, steam-reforming technologies. Even though these can be carried out at atmospheric pressure, they require temperatures of around 800 °C to be effective with steam–carbon molar ratios typically of 5 and even lower². Alcohols offer a lower-temperature option; vapour-phase steam reforming of those can be effectively carried out at temperatures of around 300 °C (ref. 3).

But does the proof of concept reported by Cortright *et al.*¹ hold the promise of an aqueous-phase technology for producing H₂ fuel from renewable biomass? To answer that question requires an interlinking of science, engineering and the economics of H₂ production.

Today's benchmark in H₂ production is provided by catalytic steam-reforming technology that uses simple hydrocarbons (such as methane and liquid petroleum gases) as feedstocks, and catalysts that are variations of well established nickel-based preparations and whose robustness guarantees operation over thousands of hours. The

Oceanography

Crossing the highway

The powerful current of the Gulf Stream is like a highway, carrying warm tropical waters from the Caribbean to Europe. The current is known to meander and to shed rotating rings of water on both sides. However, its interaction with the surrounding water tends to be

limited to the outer edges of the current, especially along the well defined northern wall of the Gulf Stream.

But satellite images presented by Xiaofeng Li and his colleagues in *Geophysical Research Letters* show a parcel of cold water from a region

known as the Middle Atlantic Bight breaching the northern boundary of the Gulf Stream, and traversing the full width of the current (*Geophys. Res. Lett.* **29**, 10.1029/2002GL015378; 2002).

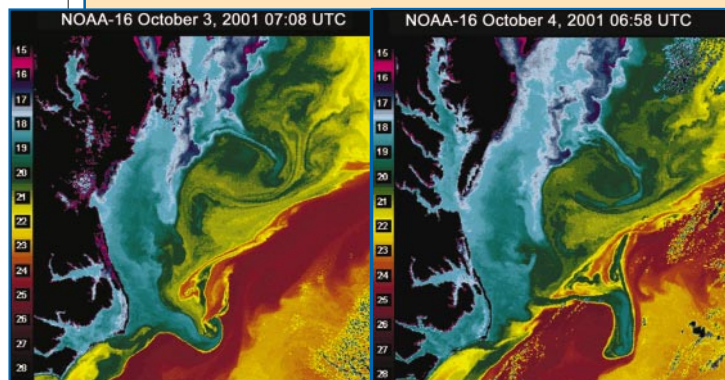
The left panel of the picture, taken at 7:08 a.m. on 3 October 2001, shows the penetration of cold water as a green tongue that extends into the main current just east of Cape Hatteras, where the Gulf Stream leaves the American East Coast and veers off into the North Atlantic Ocean. The right panel shows the fate of the intrusion about 24 hours later: the cold tongue has been swept along with the current while extending southeastwards.

In early October 2001, strong and persistent winds from northerly directions blowing along the shore

north of Cape Hatteras piled up cold water from the Middle Atlantic Bight in the corner formed by the coastline and the Gulf Stream's north wall. Under less extraordinary wind conditions, long streaks of the relatively cold shelf water are slowly mixed into the Gulf Stream along its northwestern edge. But after three days of wind speeds exceeding 12.8 m s⁻¹ — conditions unique for early autumn in the 11-year period from 1991 to 2001 — the cold coastal water broke into the main Gulf Stream and eventually crossed it.

Li and colleagues say they are not aware of any other reports of such a breaching event. After all, crossing a busy highway is rarely attempted and is even less often successful.

Heike Langenberg



conversion takes place in fractions of a second, resulting in H_2 productivities of, typically, 200 g of H_2 per kilogram of catalyst per hour². Productivity at this level means that the production costs of H_2 (expressed as the cost in US dollars per joule of calorific value contained in the molecule) are two to three times higher than for natural gas^{4,5}. However, taking into account the transportation and distribution costs of natural gas, and how pure you want the H_2 to be, the comparative cost of H_2 production can vary.

Biomass, the product of photosynthesis, could become a feedstock of choice for H_2 production, even in the absence of environmental premiums. At present, two strategies are being intensively pursued. The first is fractionation, which leads to the isolation and depolymerization of constitutive biomass fractions. An advantage of this approach is that the production of marketable goods (such as chemicals and fibres) can be integrated with the conversion of particular residual fractions to H_2 , as well as to other biofuels or to bioenergy⁶. In fact, in the absence of subsidies or carbon credits (through which countries can trade off carbon-producing industry against carbon-absorbing activities, such as the planting of crops), co-products from biomass are necessary to make bioenergy or biofuels economically viable⁷.

The second approach is through biosyntheses that produce H_2 directly. This requires photosynthesis to be conducted, as in algae, either with the removal of oxygen (to avoid poisoning the natural hydrogenases), or with oxygen-tolerant hydrogenases, which would need genetic engineering⁸. Alternatively, certain wet biomass fractions can be fermented by hydrogen-producing microorganisms in anaerobic environments where methanogenic pathways are suppressed⁹. Efforts in all of these directions are being pursued actively in academic and industrial laboratories.

Realistically, the introduction of a clean fuel such as H_2 is most likely to be achieved if the production technology can be integrated within the existing infrastructure used for natural-gas reforming. To use biomass as the source of H_2 , several objectives must be achieved. To begin with, reaction and reactor productivities must match those achieved in hydrocarbon reforming. According to the results of Cortright *et al.*¹, this is almost achieved with glycerol as substrate, but in the case of sugars as substrate much lower water–sugar ratios for aqueous-phase reforming will need to be used. Heat inputs to balance the endothermic reforming reaction must also be equivalent to those used in hydrocarbon reforming. In steam reforming, a steam–carbon molar ratio of between 3 and 5 is normally used, so for aqueous-phase reforming, an energy-equivalent liquid–water–carbon molar ratio of between 7.5 and 12.5 is needed, at an operating temperature of 260 °C.

Practical feedstocks should be chosen, such as sugar-containing hydrolysates (and not just high-purity, model sugars) or glycerol-containing liquors derived from residual fats with minimum purification. These feedstocks should preferably be obtained from high-productivity biomass crops, with little or no use of synthetic fertilizers. The biomass should also be able to yield co-products that will improve the economics of the process. Crops such as sugar cane (in tropical climates), as well as switchgrass and hybrid poplar (in temperate climates), might well be suitable.

Biofuels are becoming a viable component of tomorrow's energy mix. If the appropriate feedstocks are considered, biomass refineries seem uniquely suited to providing a variety of products — food, fibres and chemicals as well as biofuels — to satisfy society's needs in a sustainable manner. ■

Esteban Chornet and Stefan Czernik are at the National Renewable Energy Laboratory, Golden, Colorado 80401, USA. Esteban Chornet is also in the Department of Chemical Engineering, Université de Sherbrooke, Québec J1L 1H3, Canada. e-mail: esteban_chornet@nrel.gov

1. Cortright, R. D., Davda, R. R. & Dumesic, J. A. *Nature* **418**, 964–967 (2002).
2. Ridler, D. E. & Twigg, M. V. in *Catalysis Handbook* (ed. Twigg, M. V.) 225–282 (Wolfe, London, 1989).
3. Takahashi, K., Takezawa, N. & Kobayashi, H. *Appl. Catal.* **2**, 363–366 (1982).
4. Kirk-Othmer *Encyclopedia of Chemical Technology* 4th edn, Vol. 13 (Wiley, New York, 1991).
5. Blok, K., Williams, R., Katofsky, R. & Hendriks, C. *Int. J. Hydrogen Energy* **22**, 161–168 (1997).
6. Czernik, S., French, R., Feik, C. & Chornet, E. *I&EC Res.* (in the press).
7. Mann, M., Spath, P. & Kadam, K. in *Proc. 1996 US DOE Hydrog. Prog. Rev.* 249–272 (1996); Preprint NREL/CP-430-21968.
8. Ghirardi, M. L. *et al. Trends Biotechnol.* **18**, 506–511 (2000).
9. Odom, J. & Wall, J. *Appl. Environ. Microbiol.* **65**, 1300–1305 (1983).

Cognitive neuroscience

The molecules of forgetfulness

Alcino J. Silva and Sheena A. Josselyn

Not everything that we learn is useful, so the brain needs a mechanism to prevent itself being burdened by unhelpful details. The molecular details of this mechanism are now being uncovered.

Studies of the molecular and cellular foundations of cognitive processes have come of age with the development of techniques that allow genes to be over-expressed, deleted or modified in mice. These altered animals have been studied from a variety of aspects simultaneously by molecular biologists, neurophysiologists and psychologists. The result is the birth of a field that is unravelling the basis of learning, remembering¹, and now — as a paper in this issue shows — forgetting. On page 970, Genoux and colleagues² report that an enzyme known as protein phosphatase 1 (PP1) actively suppresses memories in mice, both during and after a learning exercise.

Like other biological processes, memory is regulated by yin-and-yang-like interactions between molecules with opposing functions — in this case, protein phosphatases and kinases, which respectively remove and add phosphate groups on target proteins, thereby altering their properties.

To examine the role of these interactions in forming and maintaining memories, Genoux *et al.*² generated mice that express a natural inhibitor of PP1, called inhibitor 1. When phosphorylated, this protein binds to PP1 and prevents it from working. The authors regulated the expression of inhibitor 1 in mice by using the reverse tetracycline transactivator system³: simply feeding the mice a tetracycline-like compound switches

on the gene encoding the inhibitor, enabling PP1 activity to be blocked at will.

To test the animals' memory, the authors trained them in an object-recognition task that takes advantage of the natural propensity of mice and other rodents to investigate new objects more avidly than familiar ones. Their memory for objects depends on the hippocampus — a brain structure that shows robust expression of the inhibitor.

The authors' initial results help to explain one of the earliest discoveries of modern experimental psychology. In landmark experiments published in 1885, Hermann Ebbinghaus⁴ showed that distributing training into several sessions results in stronger memories than equivalent amounts of training crammed into a single session (students take note!). Genoux *et al.* now find that massed training triggers more PP1 activity than distributed training. Remarkably, when the authors switched on inhibitor 1 specifically during massed training, blocking the PP1 activity, this type of training became as effective as distributed training. So PP1 apparently suppresses memory formation during massed training — but how?

One of PP1's targets is a gene-transcription factor called CREB, which becomes inactive when dephosphorylated by PP1. When phosphorylated, CREB directs the transcription of genes with specific CREB-binding DNA sequences in their control regions⁵, and so is needed for proteins to be

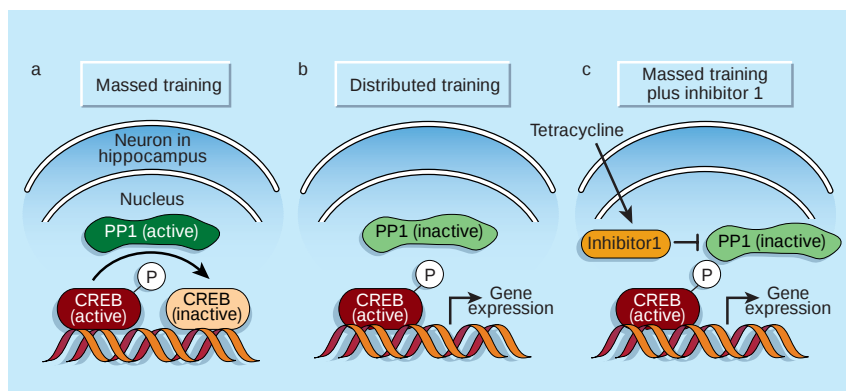


Figure 1 Learning and forgetting — a tale of molecular antagonism. Genoux *et al.*² studied the molecular and cellular basis of memory by inserting into mice a gene that is activated, particularly in the brain's hippocampus, by feeding the mice the drug tetracycline. The protein produced from this gene, inhibitor 1, represses an enzyme, protein phosphatase 1 (PP1). a, When the mice were trained in a certain learning task in the absence of tetracycline, with the training crammed into one session, PP1 was active, and dephosphorylated (thereby inactivating) the gene-transcription factor CREB (a circled 'P' represents phosphorylation). This blocked the expression of genes needed for long-term memory formation. b, If training was distributed among more sessions, PP1 activity was lower, CREB was active, and the appropriate genes could be expressed. c, If inhibitor 1 was induced during massed training, repressing PP1, the levels of active CREB rose to those seen during distributed training in b.

produced from these genes. Since the late 1960s, it has been known that blocking protein synthesis results in memory deficits that emerge within hours of training⁶. Later studies of species ranging from the marine mollusc *Aplysia* to flies, mice and rats have put this finding on a more detailed molecular footing, showing that CREB is required for the long-term processes involved in memory formation⁷.

It now appears that one of the ways in which inhibitor 1 enhances memory during massed training is by blocking the PP1-dependent dephosphorylation of CREB that would normally occur (Fig. 1). That alters CREB's activity, which Genoux *et al.* studied by looking at the CREB-regulated expression of an inserted gene encoding β -galactosidase — an enzyme that produces a characteristic blue colour from its substrate. CREB activity (as judged by β -galactosidase expression) was higher after distributed training than massed training when inhibitor 1 was not switched on. But when activated, the inhibitor abolished this difference, increasing CREB function during massed training to levels comparable to those during distributed training.

Earlier work^{8–10} showed that genetic manipulation of CREB also affects the learning regimen required to form long-term memories. In general, decreasing CREB levels increases the numbers of trials and the length of inter-trial intervals required for such memories to be formed, whereas increasing CREB levels has the opposite effect. Genoux *et al.*'s results suggest that the balance between the levels of active CREB and PP1 during training determines whether long-term memories are laid down.

The authors also found that inhibiting

PP1 had almost immediate effects on memory formation. But these effects are unlikely to involve CREB-dependent transcription, which is an inherently slow process. Instead, these results point to molecules implicated in the early stages of memory formation, such as calcium/calmodulin-dependent protein kinase II (CaMKII)¹¹. During learning, this enzyme becomes activated by increases in calcium levels inside neurons, with the result that several critical signalling molecules at neuronal junctions (synapses) are phosphorylated, and CaMKII itself becomes constitutively active. Such events are thought to be central to the changes in neuronal communication required for learning and memory. Indeed, genetic and pharmacological manipulations that interfere with CaMKII impair learning¹¹. Genoux *et al.* now show that inhibiting PP1 results in higher levels of activated CaMKII. So, during massed training PP1 suppresses both short-term and long-term processes involved in learning and memory formation.

Can this phosphatase also suppress memories after they have formed? Learning is thought to trigger molecular changes (short- and long-term) that facilitate neuronal communication and are important for memory. Just as digital information is encoded by altering the grooves in a compact disc during burning, so memories are encoded by altering synapses in the brain during training. By means not yet understood, neurons can decode these changes and use them to reconstruct the learned information. Unlike a compact disc, however, the brain seems to store information in a highly dynamic manner: psychological studies suggest that

the brain cares more about the usefulness of stored memories than their fidelity. Most of what we perceive and process is quickly discarded, and what we do remember tends to fade and change with time. There is a good reason for this: flawless recall would burden the brain with useless details, perhaps at the expense of storing and recalling information that is useful.

Genoux and colleagues have found that PP1 is involved in forgetting. The authors trained mice to find a submerged platform in a round pool of opaque water — another task that requires the hippocampus. To test the animals' memory, Genoux *et al.* then removed the platform. The mice initially searched for the platform in the correct quadrant; but with time, memory faded and performance declined. The authors found that blocking PP1 activity during training accelerated learning, but did not affect memory decline. But impairing PP1 activity after learning all but halted the memory decline for four to six weeks. Interestingly, studies¹² of an inserted CaMKII gene showed that interfering with its function after training destabilizes memory. So the antagonistic interactions between PP1 and CaMKII that are seen during learning may also be important later on, in maintaining memories.

Previous neurophysiological studies have shown that PP1, CaMKII and CREB modulate the changes in synaptic strength that are thought to underlie learning and memory¹³. These results mirror Genoux *et al.*'s molecular and behavioural findings. Similar examples of convergence between different types of study are now common in work on cognition at the molecular and cellular levels, showing that we are well on our way to unravelling the biology of learning and forgetting. ■

Alcino J. Silva and Sheena A. Josselyn are in the Departments of Neurobiology, Psychology and Psychiatry and the Brain Research Institute, 695 Charles Young Drive South, Gonda Building, University of California Los Angeles, California 90095-1761, USA.

e-mail: silvaa@mednet.ucla.edu

- Silva, A. J., Smith, A. M. & Giese, K. P. *Annu. Rev. Genet.* **31**, 527–546 (1997).
- Genoux, D. *et al.* *Nature* **418**, 970–975 (2002).
- Mayford, M., Mansuy, I. M., Müller, R. U. & Kandel, E. R. *Curr. Biol.* **7**, R580–R589 (1997).
- Ebbinghaus, H. *Über das Gedächtnis. Untersuchungen zur Experimentellen Psychologie* (Duncker & Humboldt, Leipzig, 1885).
- Mayr, B. & Montminy, M. *Nature Rev. Mol. Cell Biol.* **2**, 599–609 (2001).
- Davis, H. P. & Squire, L. R. *Psychol. Bull.* **96**, 518–559 (1984).
- Silva, A. J., Kogan, J. H., Frankland, P. W. & Kida, S. in *Annual Review of Neuroscience* Vol. 21 (ed. Cowan, W. M.) 127–148 (Annual Reviews, Palo Alto, California, 1998).
- Yin, J., Del Vecchio, M., Zhou, H. & Tully, T. *Cell* **81**, 107–115 (1995).
- Josselyn, S. A. *et al.* *J. Neurosci.* **21**, 2404–2412 (2001).
- Kogan, J. H. *et al.* *Curr. Biol.* **7**, 1–11 (1997).
- Lisman, J., Schulman, H. & Cline, H. *Nature Rev. Neurosci.* **3**, 175–190 (2002).
- Mayford, M. *et al.* *Science* **274**, 1678–1683 (1996).
- Blitzer, R. D. *et al.* *Science* **280**, 1940–1942 (1998).

Widespread local house-sparrow extinctions

Agricultural intensification is blamed for the plummeting populations of these birds.

House-sparrow populations have declined sharply in Western Europe in recent decades^{1,2}, but the reasons for this decline have yet to be identified, despite intense public interest in the matter. Here we use a combination of field experimentation, genetic analysis and demographic data to show that a reduction in winter food supply caused by agricultural intensification is probably the principal explanation for the widespread local extinctions of rural house-sparrow populations in southern England. We show that farmland populations exhibit fine-level genetic structuring and that some populations are unable to sustain themselves (sinks), whereas others act as sources.

In rural England, house sparrows (*Passer domesticus*) tend to occur in localized populations aggregated around farmyards³. We used standard techniques of nest recording, mark resighting and microsatellite-based molecular genetics to evaluate annual productivity, survival rates and gene flow in four populations on lowland farmland in Oxfordshire, UK. We also provided these sparrows with supplementary food *ad libitum* during winter to investigate the effect of food supply on their survival.

The minimum distance between the four populations (designated A–D) was 6 km; the maximum distance was 24 km. Population A was chosen because of the availability of historical records, whereas populations B–D were selected at random. All four populations were located on farms that are typical of the mixed pastoral and arable agricultural system that prevails in Oxfordshire; population sizes during the breeding season were about 35, 150, 40 and 60 birds, respectively.

For population A, estimates of population size from historical records reveal an 80% decline over the past 30 years (150 breeding individuals in 1971; 35 in 2000). Although equivalent historical data for populations B–D are unavailable, landowners reported that their populations had been largely stable over the previous decade.

Consistent with this, genetic analysis provided evidence of a genetic ‘bottleneck’, as indicated by a significant departure from mutation-drift equilibrium (that is, excess heterozygosity)⁴ at five microsatellite loci ($P < 0.05$ under the assumption of a two-phased mutation model), only in population A. Population-genetic analyses revealed low but significant differentiation across the four populations ($F_{ST} = 0.011$, $P < 0.001$) and in all pairwise comparisons. Differentiation at such a small spatial scale is rare

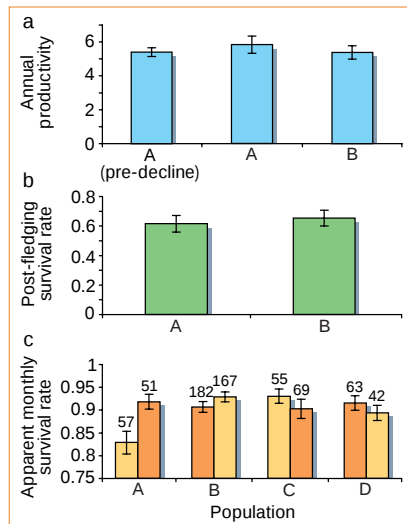


Figure 1 Annual productivity, survival rates and effect of supplementary feeding on rural house-sparrow populations. **a**, Comparison of annual productivity (calculated as the product of the mean number of breeding attempts per female (n , 35–138), the mean number fledged per successful attempt (n , 76–266) and the nest survival rate over the entire nesting period (n , 92–370)) in populations A and B (1998–2001), and in population A before the decline (1960–71). None of the comparisons was significant (means $\pm$ s.e. are shown). **b**, Post-fledging survival rates (defined as the period between fledging and nutritional independence) in populations A (n = 20 broods) and B (n = 23) were not significantly different (means $\pm$ s.e. are shown). **c**, Supplementary feeding significantly increased apparent monthly survival rate only in population A ($\chi^2 = 10.28$, d.f. = 1, $P = 0.001$). Survival rates include an unknown component of emigration (hence termed ‘apparent’). Orange bars represent years in which supplementary food was provided; yellow bars represent control years. After regular overwinter ringing, the proportion of colour-ringed individuals increased during each winter at all four sites, suggesting a low and consistent population turnover. Population A was randomly paired with C (control sites in winter 1998–99, fed in winter 1999–2000) and B with D (fed in 1999, control in 2000). Opposing trends in populations A and C preclude an effect of a particular year. Values are means $\pm$ s.e.; sample sizes are shown above each bar and include both adults and young.

in birds⁵ and is likely to reflect the sedentary nature of the house sparrow⁶ and/or the effects of population decline and habitat fragmentation.

Demographic data revealed that populations A and B (for which directly comparable data were available) did not differ in either annual productivity or post-fledging survival rate (Fig. 1a, b). Similarly, annual productivity for population A was the same before the decline (1967–71) as for the present day (Fig. 1a).

However, apparent monthly overwinter survival rates (that is, survival rate plus an unknown component of emigration) did

vary between the four populations, with that of population A being markedly lower than the others (Fig. 1c). Provision of supplementary seed food in winter increased overwinter survival of population A only (Fig. 1c); the apparent survival rate over the entire November–March period increased from 0.391 ± 0.059 (s.e.) in 1998–99 to 0.653 ± 0.058 in response to supplementary feeding in 1999–2000.

Modelling these demographic data as a simple Leslie matrix to give a qualitative measure of population change (we were unable to measure density dependence) revealed a marked decline in population A. This contrasts with populations B–D, which all increased.

We conclude that population A is limited by overwinter food supply and, with restricted movement between populations, immigration may be insufficient to sustain it in the future. Key changes resulting from agricultural intensification that are likely to have reduced winter food supplies for house sparrows in recent decades are the switch from spring sowing of cereals to autumn⁷ and an increase in bird-proof storage of grain and other animal-feed stocks⁸.

More generally, our results indicate that the decline of the house sparrow in rural areas of England may have arisen as a result of local extinctions. Shortage of food on some farms in winter reduces survival, and there may be insufficient movement of sparrows between farms to provide enough recruits to sink populations from source ones. The results of a postal survey of a random selection of farmers in Britain seem to support this suggestion, revealing a 10% extinction rate for farmland populations in areas of England outside Oxfordshire (39% declined, 6% increased; $n = 113$) and a 20% extinction rate in Oxfordshire (45% declined, 1% increased; $n = 104$) over the past 20 years.

We contend that local extinction without recolonization may eventually reduce migration between populations as a result of greater separation between remaining populations, a compounding effect that could pervade the rural landscape and which has been predicted by classical metapopulation theory⁹.

David G. Hole*, Mark J. Whittingham*, Richard B. Bradbury†, Guy Q. A. Anderson†, Patricia L. M. Lee‡, Jeremy D. Wilson§, John R. Krebs*

*Farmland Bird Group, Edward Grey Institute of Field Ornithology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK
e-mail: david.hole@linacre.ox.ac.uk

Royal Society for the Protection of Birds, †The Lodge, Sandy, Bedfordshire SG19 2DL, UK, and §Dunedin House, Edinburgh EH4 3TP, UK
‡School of Biological Sciences, University of Wales, Swansea, Singleton Park, Swansea SA2 8PP, UK

1. Krebs, J. R., Wilson, J. D., Bradbury, R. B. & Siriwardena, G. M. *Nature* **400**, 611–612 (1999).
2. Gregory, R. D. et al. *The State of the UK's Birds 2000* (RSPB, BTO and WWT, Sandy, UK, 2001).

Ageing

Cognitive change and the APOE ε4 allele

There is a marked variation in whether people retain sufficient cognitive function to maintain their quality of life and independence in old age, even among those without dementia, so it would be valuable to identify the determinants of normal age-related cognitive change^{1,2}. We have retested non-demented 80-year-olds who were participants in the Scottish Mental Survey of 1932, and find that the variation in their non-pathological cognitive change from age 11 to 80 is related to their apolipoprotein E (APOE) genotype. This effect of the APOE ε4 allele on normal cognitive ageing may be mediated by a mechanism that is at least partly independent of its predisposing effect towards Alzheimer's disease.

The Scottish Mental Survey of 1932 tested cognitive ability in almost everyone attending school in Scotland in June 1932 and who was born in 1921 ($n = 87,498$; ref. 3). From the Lothian birth cohort of 1921, 491 surviving participants (206 men) of the 1932 survey had their Moray House Test (MHT) score at age 11 traced. All lived independently in the community. Disease history and current medication were established at interview.

Subjects with Mini-Mental State Examination (MMSE) scores that were suggestive of dementia (score < 24; 4 men, 4 women), or with any history of dementia-related illness (3 men, 2 women), were omitted from the study. There were 466 subjects (190 men) for whom we had Moray House Test scores at both age 11 and age 80, and whose APOE genotype we were able to determine.

The MHT takes 45 minutes and comprises 71 verbal and non-verbal reasoning questions. It has criterion validity in youth and old age⁴. Up to 50% of the variation between subjects in the MHT remains stable from age 11 to about 80 (ref. 4). Subjects took the test in 1932 and again in 1999–2001. Scores at each age were controlled for age (in days) and were converted to IQ-type scores (mean, 100; s.d., 15). Subjects were classified according to the presence ($n = 121$) or absence ($n = 345$) of

3. Summers-Smith, J. D. *The Sparrows* (Poyser, Calton, 1988).
4. Cornuet, J. M. & Luikart, G. *Genetics* **144**, 2001–2014 (1996).
5. Crochet, P. A. *Mol. Ecol.* **9**, 1463–1469 (2000).
6. Wernham, C. V. et al. *The Migration Atlas: Movements of the Birds of Britain and Ireland* (Poyser, London, in the press).
7. Chamberlain, D. E., Fuller, R. J., Bunce, R. G. H., Duckworth, J. C. & Shrubbs, M. J. *Appl. Ecol.* **37**, 771–788 (2000).
8. Robinson, R. A. & Sutherland, W. J. *J. Appl. Ecol.* **39**, 157–176 (2002).
9. Hanski, I. *Nature* **396**, 41–49 (1998).

Competing financial interests: declared none.

one or more APOE ε4 allele(s)⁵.

At age 11, MHT scores (mean, s.d.) were similar ($t = 0.91$, $P = 0.36$) for those with the ε4 allele (99.4, 15.2) and for those without it (100.8, 14.4). But at age about 80, MHT scores were significantly different ($t = 2.64$, $P = 0.009$) between the two groups (with the ε4 allele, 97.0, 15.7; without the ε4 allele, 101.1, 14.2).

We used analysis of covariance to examine the significance and effect sizes of contributors to cognitive change from age 11 to 80. Effects of age-11 MHT score ($F_{1,461} = 334.8$, $P < 0.001$, $\eta^2 = 0.42$; Pearson's correlation between MHT at age 11 and at age 80, 0.65, $P < 0.001$), sex ($F_{1,461} = 8.7$, $P = 0.003$, $\eta^2 = 0.018$) and presence of the ε4 allele ($F_{1,461} = 5.2$, $P = 0.02$, $\eta^2 = 0.011$) were significant; sex and ε4 status did not interact. Cardiovascular disease history was not significantly associated with cognitive function. People with and without the APOE ε4 allele did not differ in the number and type of medications that they took.

We reran these models to investigate the possibility that the effects of ε4 might be caused by incipient dementia in some APOE ε4⁺ individuals. Including only those subjects with MMSE scores of ≤ 28 resulted in an increased effect of ε4 ($F_{1,342} = 8.4$, $P = 0.004$, $\eta^2 = 0.024$). The effect was also evident after excluding those subjects whose IQ decrement from age 11 to 80 was greater than two standard deviations ($F_{1,459} = 5.1$, $P = 0.02$, $\eta^2 = 0.011$).

This study has the advantage of examining the same individuals with the same cognitive test in childhood and in old age. Individual differences in childhood IQ are the most reliable predictor of late-life cognitive performance (accounting for at least 42% of variance). Possession of APOE ε4 is unrelated to differences in mental ability in youth, but is significantly associated with mental ability in old age and the change in ability score from youth.

Incipient dementia may have been more common in our APOE ε4-positive individuals and so could account for part of the effect we have observed. But we contend that this is not the only possible explanation, because the effect of APOE ε4 increases when the MMSE threshold is raised to include only near-perfect scorers,

and remains similar after excluding those whose IQ decrement from age 11 to 80 was greater than 2 s.d. Moreover, the number of incipient cases expected in this age group^{6,7}, and the effect size of APOE on Alzheimer's disease⁸, mean that each case would require a massive cognitive decline to account for the observed effects, and no such skewing or bimodal distribution was seen in either ε4 subgroup. The results do not seem to be due to selective attrition, because the proportion of people with ε4 alleles is similar to that in other, younger groups from Scotland⁹.

Further clinical and pathological investigation is required to determine whether Alzheimer's disease or other changes could be responsible for the differences we observe. APOE isoforms show differences in their binding to receptors, lipoproteins and amyloid-β, in atherosclerosis, neurite extension, neuroprotection and repair¹⁰. These various interactions could influence cognitive ageing independently of those that operate in Alzheimer's disease, which remain uncertain¹⁰.

Identifying a factor that influences non-pathological, lifetime cognitive change has large public-health implications because, in the absence of a sharp risk threshold, most adverse events — and most of the personal and economic burdens that these bring — occur to old people who are within the 'normal' part of the distribution¹¹.

Ian J. Deary*, Martha C. Whiteman*, Alison Pattie*, John M. Starr†, Caroline Hayward‡, Alan F. Wright‡, Andrew Carothers‡, Lawrence J. Whalley§

*Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK
e-mail: i.deary@ed.ac.uk

†Department of Geriatric Medicine, Royal Victoria Hospital, Craigleith Road, Edinburgh EH4 2DN, UK

‡Medical Research Council Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

§Department of Mental Health, University of Aberdeen, Clinical Research Centre, Royal Cornhill Hospital, Cornhill Road, Aberdeen AB25 2ZH, UK

1. US National Research Council *The Aging Mind* (Nat'l Acad. Press, Washington DC, 2000).
2. Anstey, K. & Christensen, H. *Gerontology* **46**, 163–177 (2000).
3. Scottish Council for Research in Education *The Intelligence of Scottish Children: A National Survey of an Age Group* (Univ. London Press, 1933).
4. Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R. & Starr, J. M. *Intelligence* **28**, 49–55 (2000).
5. Wernham, P. R., Price, W. P. & Blundell, G. *Lancet* **33**, 1158–1159 (1991).
6. Breitner, J. C. et al. *Neurology* **55**, 161–162 (2000).
7. Corder, E. H. et al. *Nature Genet.* **7**, 180–184 (1994).
8. Slooter, A. J. et al. *Arch. Neurol.* **55**, 964–968 (1998).
9. Robertson, F. W. & Cumming, A. M. *Arteriosclerosis* **5**, 283–292 (1985).
10. Mahley, R. W. & Rall, S. C. Jr *Annu. Rev. Genom. Hum. Genet.* **1**, 507–537 (2000).
11. Rose, G. *The Strategy of Preventive Medicine* (Oxford Univ. Press, 1992).

Competing financial interests: declared none.

Nuclear fusion

Fast heating scalable to laser fusion ignition

Rapid heating of a compressed fusion fuel by a short-duration laser pulse is a promising route to generating energy by nuclear fusion¹, and has been demonstrated on an experimental scale using a novel fast-ignitor geometry². Here we describe a refinement of this system in which a much more powerful, pulsed petawatt (10^{15} watts) laser creates a fast-heated core plasma that is scalable to full-scale ignition, significantly increasing the number of fusion events while still maintaining high heating efficiency at these substantially higher laser energies. Our findings bring us a step closer to realizing the production of relatively inexpensive, full-scale fast-ignition laser facilities.

In advanced laser ignition of fusion^{3,4}, high-density energetic electrons generated by petawatt lasers instantaneously heat a compressed fusion fuel to its ignition temperature with high coupling efficiency⁵. We tested fast heating by a petawatt laser⁶ and the GEKKO XII laser systems on targets in which a hollow gold cone (30° or 60° angle) was inserted into a deuterated polystyrene ('CD') shell (500 μm diameter, 7 μm thick)².

The shell was imploded using nine beams of the laser system operated at a wavelength of 0.53 μm and with an energy of 2.5 kJ for 1.2-ns flat-top pulses. The fast-heating laser was injected into the cone's interior and generated energetic electrons at the end of the cone, at the stagnation of the shell compression with a power of 0.5 petawatts (PW). The imploded core plasma was created near to the centre of the shell, close to the tip of the cone; the compressed density was estimated by using an X-ray backlight method² as $50\text{--}100\text{ g ml}^{-1}$ for cores of diameter 30–50 μm . A single laser oscillator⁷ was used for both laser systems to provide perfect synchronization between shell compression and fast electron heating.

To quantify the heating of these highly compressed plasmas, we measured the increase in thermonuclear neutron production as a function of the injection timing of the heating pulse, with respect to the time of peak compression. Neutrons are generated by fusion of two deuterium nuclei to form a helium nucleus (atomic configuration, $d(d,n)^3\text{He}$) in the compressed CD plasma.

Neutron energy spectra were obtained using time-of-flight scintillator/photomultiplier detectors. The coincidence of signals from detectors at different distances and angles confirmed that the neutrons were thermonuclear in origin. Neutron enhancement was about three orders of magnitude

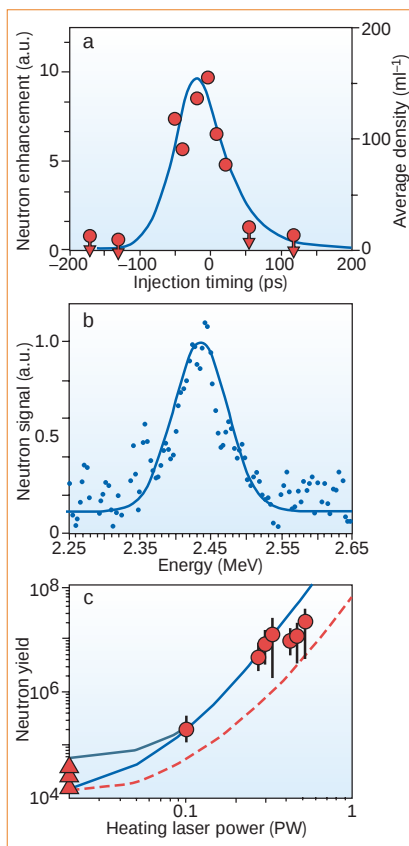


Figure 1 Fast heating of highly compressed plasmas with a petawatt (PW)-class laser pulse. **a**, Neutron enhancement (a.u., arbitrary units) as a function of injection timing (in picoseconds) of the heating pulse. Line represents the temporal profile of the average density of the compressed plasma from a hydrodynamic simulation. Arrows denote data points that fall below the background noise level ($S/N < 1$) or that correspond to the noise level. **b**, Neutron spectrum observed from an angle of 42° for a 0.5-PW laser heating at maximum compression. Line is a Gaussian fit to the data points, indicating an ion temperature of 0.8 keV. **c**, Yield of thermonuclear neutrons from the compressed plasma as a function of heating laser power for the pulse duration of 0.6 ps. Solid and dashed lines, neutron yields for an energy coupling from laser to core of 30% and 15%, respectively; neutron yields obtained without a heating pulse are shown in the 0.02-PW position.

at 0.5 PW, compared with neutrons obtained with no heating pulse ($2\text{--}5 \times 10^4$ for a 1.2 flat-pulse implosion).

Figure 1a shows this enhancement as a function of injection timing of the heating pulse. The timing of heating was checked with X-ray streak images, as well as the injection timing of the pulse to maximum compression, from hydrodynamic simulations of the shell implosion. Enhancement was evident during ± 40 ps, which corresponds to the stagnation time of the imploded plasma; the heating pulse is 0.6 ps, which is two orders of magnitude shorter than the stagnation time. These results indicate that heating for ignition might be achieved by using pulses that are close to the duration of stagnation.

If a gaussian profile is fitted to the neutron

spectrum (Fig. 1b), the spectral width of the 0.5-PW heating shot is 90 ± 5 keV, which is greater than that (60 keV) for 0.1-PW heating. The width of the 0.1-PW-heating spectrum was similar to that for no heating pulse, which was less than, or close to, the spectral resolution corresponding to the ion temperature of 0.4 keV. Taking into account the spectral resolution, the width (90 ± 5 keV) for 0.5-PW heating corresponds to an ion temperature of 0.8 ± 0.1 keV, indicating that the temperature of core plasmas could be doubled by this heating.

This finding is consistent with the enhancement, by three orders of magnitude, of neutron yield through heating; this indicates that the temperature increases from 0.3–0.4 to 0.8 keV. Our results are also consistent with the change in intensity and spectra of X-rays from heated core plasmas. The intensity increases by a factor of 1.5–2.0 compared with the absence of a heating pulse, and a continuum slope of the X-ray spectra (3–4 keV), temporally resolved with an X-ray streak camera, shows that the increase in temperature (1 ± 0.1 keV compared with 0.4 keV) is more than doubled.

Neutron yields are summarized in Fig. 1c for 0.6-ps laser pulses. Simple predictions of the neutron yield normalized to the yield without heating from energy conservation are also shown as a function of the heated laser energy, for the coupling from laser to the core plasma of 15% and 30%. The yield increases with the energy of the heating laser, implying almost constant coupling from the laser to the core plasma. However, there may be a small decrease in the coupling, from about 30% to 20%, as the laser power is increased from 0.1 PW to 0.5 PW. This could be due to an increase in the energetic electron temperature, resulting in a reduction in the stopping power of electrons for a fixed spot diameter.

Efficient fast heating of imploded plasmas has been accomplished with a petawatt laser at powers that are almost equivalent to those required in fast-ignition conditions. The period for sufficient heating is similar to the stagnation time (± 40 ps), suggesting that the heating laser's energy could be increased to ignite the fuel with a heating pulse of up to 10–20 ps or more at similar irradiance. It may eventually be possible to ignite a compressed deuterium–tritium fusion plasma with a relatively inexpensive fast-ignition facility comprising a petawatt-class laser.

R. Kodama & the Fast-Ignitor Consortium*

Institute of Laser Engineering, Osaka University, 2-6 Yamada-oka, Suita Osaka 565-0871, Japan e-mail: ryo@ile.osaka-u.ac.jp

*H. Shiraga, K. Shigemori, Y. Toyama, S. Fujioka, H. Azechi, H. Fujita, H. Habara†, T. Hall‡, Y. Izawa, T. Jitsuno, Y. Kitagawa, K. M. Krushelnick§, K. L. Lancaster†§, K. Mima, K. Nagai, M. Nakai, H. Nishimura, T. Norimatsu,

P. A. Norreys†, S. Sakabe ||, K. A. Tanaka ||, A. Youssef, M. Zepf¶, T. Yamanaka †Rutherford Appleton Laboratory, Oxfordshire OX11 0QX, UK; ‡University of Essex, Colchester CO4 3SQ, UK; §Blacket Laboratory, Imperial College, London SW7 2BZ, UK; ||Institute of Laser Engineering and Faculty of Engineering, Osaka University, Suita Osaka 565-0871, Japan; ¶Queen's University of Belfast, Belfast BT7 1NN, UK

1. Tabak, M. *et al. Phys. Plasmas* **1**, 1626–1634 (1994).
2. Kodama, R. *et al. Nature* **412**, 798–802 (2001).
3. Norreys, P. *et al. Phys. Plasmas* **7**, 3721–3726 (2000).
4. Kodama, R. *et al. Phys. Plasmas* **8**, 2268–2274 (2001).
5. Key, M. H. *Nature* **412**, 775–776 (2001).
6. Kitagawa, Y. *et al. Phys. Plasmas* **9**, 2202–2207 (2002).
7. Yoshida, H. *et al. Conf. Laser Elec. Opt.* **2002**, 4, 402–403 (2002).
Competing financial interests: declared none.

Tumorigenesis

RAF/RAS oncogenes and mismatch-repair status

Genes of the *RAF* family encode kinases that are regulated by Ras and mediate cellular responses to growth signals. Activating mutations in one *RAF* gene, *BRAF*, have been found in a high proportion of melanomas and in a small fraction of other cancers¹. Here we show that *BRAF* mutations in colorectal cancers occur only in tumours that do not carry mutations in a *RAS* gene known as *KRAS*, and that *BRAF* mutation is linked to the proficiency of these tumours in repairing mismatched bases in DNA. Our results not only provide genetic support for the idea that mutations in *BRAF* and *KRAS* exert equivalent effects in tumorigenesis², but also emphasize the role of repair processes in establishing the mutation spectra that underpin human cancer.

To determine how alterations in *BRAF* and *KRAS* might affect one another, we systematically evaluated mutations in these genes in 330 colorectal tumours (Table 1).

We identified 32 mutations in *BRAF*: 28 cases with thymine-to-adenine (T–A) transversions at nucleotide position 1,796 (corresponding to an amino-acid swap of glutamate for valine at residue 599; V599E), and one case each of a guanine-to-thymine (G–T) transversion at nucleotide 1,382 (R461I), a T–G transversion at nucleotide 1,385 (I462S), a G–A transition at nucleotide 1,388 (G463E), and an A–G transition at nucleotide 1,798 (K600E). All but two of these mutations seemed to be heterozygous, and in all 20 cases for which normal tissue was available, the mutations were shown to be somatic. In the same set of tumours, there were 169 mutations in *KRAS*, including alterations to codons 12, 13, 59 and 61. No tumour exhibited mutations in both *BRAF* and *KRAS*.

Mutations in either *BRAF* or *KRAS* occurred in all Duke's stages of cancer (results not shown) and also in premalignant lesions. Mutations in both genes seemed to be more common in adenomas larger than 1 cm across than they were in smaller adenomas.

There was also a striking difference in the frequency of *BRAF* mutations between cancers with and without mismatch-repair (MMR) deficiency ($P < 10^{-6}$, χ^2 test; Table 1). All but one of the 15 *BRAF* mutations identified in MMR-deficient cases resulted in a V599E substitution.

These results provide strong support for the hypothesis that *BRAF* and *KRAS* mutations are equivalent in their tumorigenic effects². Both genes seem to be mutated at a similar phase of tumorigenesis, after initiation but before malignant conversion. Moreover, we found no tumour that concurrently contained both *BRAF* and *KRAS* mutations. In view of the large number of mutations of both genes found in colorectal cancers, this observation is highly statistically significant ($P < 10^{-6}$, χ^2 -test) and cannot be easily explained in other ways. This conclusion

could not have been reached through the study of melanomas or of most other tumour types in which only one of the two genes is commonly mutated. It is consistent with biochemical observations³ and was suggested by Davies *et al.*¹.

Our results also show that MMR-deficient tumours have a very high incidence of *BRAF* mutations and a lower incidence of *KRAS* mutations compared with MMR-proficient colorectal cancers. This is consistent with the idea that both tumour types progress through the same biochemical pathways, but that the mutation spectrum depends on the nature of the underlying genetic instability⁴. The V599E mutation is the most frequent nucleotide substitution ever identified in a repair-deficient tumour.

The only other tumour type with a *BRAF*-mutation frequency as high as that seen in MMR-deficient colorectal cancers is melanoma¹. Melanomas and MMR-deficient colorectal cancers also share a high incidence of mutations in the oncogene that encodes β -catenin^{5,6}. It will be interesting to see whether melanomas have a repair defect that makes them susceptible to the types of mutation found in MMR-deficient colorectal cancers, and to determine what structural or sequence elements surrounding *BRAF* codon 599 make it prone to mutagenesis in a repair-deficient background.

Harith Rajagopalan, Alberto Bardelli*, Christoph Lengauer, Kenneth W. Kinzler, Bert Vogelstein, Victor E. Velculescu
Sidney Kimmel Comprehensive Cancer Center, Howard Hughes Medical Institute, and Program in Cellular and Molecular Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21231, USA
e-mail: velculescu@jhmi.edu
*On leave from the Institute for Cancer Research, University of Torino, Torino, Italy

Table 1 *BRAF* and *KRAS* mutations in colorectal tumours

Tumours	No. of cases	<i>BRAF</i> mutation	<i>KRAS</i> mutation
All types	330	32 (10%)	169 (51%)
<i>BRAF</i> mutants	1	R461I	WT
	1	I462S	WT
	1	G463E	WT
	28	V599E	WT
	1	K600E	WT
<i>KRAS</i> mutants	169	WT	MUT
Other	129	WT	WT
Clinical cancers	276	30 (11%)	154 (56%)
Adenomas > 1 cm	20	2 (10%)	12 (60%)
Adenomas ≤ 1 cm	34	0 (0%)	3 (9%)
MMR-deficient cancers	49	15 (31%)	21 (43%)
MMR-proficient cancers	227	15 (7%)	133 (59%)

DNA was purified from microdissected primary tumours ($n=54$), first-passage xenografts ($n=189$) or cell lines ($n=87$) as described⁷. The complete coding sequences of exons 11 and 15 of *BRAF* and exons 2 and 3 of *KRAS* were amplified by polymerase chain reaction using intronic primers and the products were sequenced as described⁸. Mutations were identified using the Mutation Explorer package (SoftGenetics). This strategy allowed us to identify all mutations previously known to occur in these two genes. Mismatch-repair (MMR) deficiency was assessed by analysis of microsatellite instability, using the BAT26 marker and at least 12 microsatellite repeat markers⁹. WT, wild-type sequence; MUT, mutations in codons 12, 13, 59 or 61 in *KRAS*.

1. Davies, H. *et al. Nature* **417**, 949–954 (2002).
2. Storm, S. M. & Rapp, U. R. *Toxicol. Lett.* **67**, 201–210 (1993).
3. Marais, R., Light, Y., Paterson, H. F., Mason, C. S. & Marshall, C. J. *J. Biol. Chem.* **272**, 4378–4383 (1997).
4. Kinzler, K. W. & Vogelstein, B. *Cell* **87**, 159–170 (1996).
5. Sparks, A. B., Morin, P. J., Vogelstein, B. & Kinzler, K. W. *Cancer Res.* **58**, 1130–1134 (1998).
6. Rubinfeld, B. *et al. Science* **275**, 1790–1792 (1997).
7. Thiagalingam, S. *et al. Proc. Natl Acad. Sci. USA* **98**, 2698–2702 (2001).
8. Wang, T. L. *et al. Proc. Natl Acad. Sci. USA* **99**, 3076–3080 (2002).
9. Markowitz, S. J. *Clin. Oncol.* **18** (suppl.), 75–80 (2000).

Competing financial interests: declared none.

brief communications is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on *Nature's* website (www.nature.com/nature) or on request from nature@nature.com

Coordination of circadian timing in mammals

Steven M. Reppert & David R. Weaver

Department of Neurobiology, University of Massachusetts Medical School, LRB-728, 364 Plantation Street, Worcester, Massachusetts 01605, USA

Time in the biological sense is measured by cycles that range from milliseconds to years. Circadian rhythms, which measure time on a scale of 24 h, are generated by one of the most ubiquitous and well-studied timing systems. At the core of this timing mechanism is an intricate molecular mechanism that ticks away in many different tissues throughout the body. However, these independent rhythms are tamed by a master clock in the brain, which coordinates tissue-specific rhythms according to light input it receives from the outside world.

Circadian rhythms, as exemplified by the sleep/wake cycle, are the outward manifestation of an internal timing system. The full force of genetic, molecular and biochemical approaches, complemented by precise behavioural observations, has rapidly advanced our knowledge of circadian timing in mammals. The focal point of this system is a master clock, located in the suprachiasmatic nuclei (SCN) of the anterior hypothalamus, which orchestrates the circadian programme¹. Principal advances in understanding the molecular and biochemical basis of circadian timing have provided a rapidly evolving model of the underlying 'clockwork'. Recent

developments have also revolutionized our view of SCN input and output mechanisms. These include the discovery of a new visual pathway from retina to the SCN that entrains (synchronizes) circadian rhythms to the solar day, and the elucidation of ways in which the SCN clock ultimately generates output rhythms in physiology and behaviour.

Defining the molecular basis of circadian timing in mammals has profound implications. In terms of fundamental brain mechanisms, the circadian system is among the most tractable models for providing a complete understanding of the cellular and molecular events connecting genes to behaviour. Thorough dissection of the

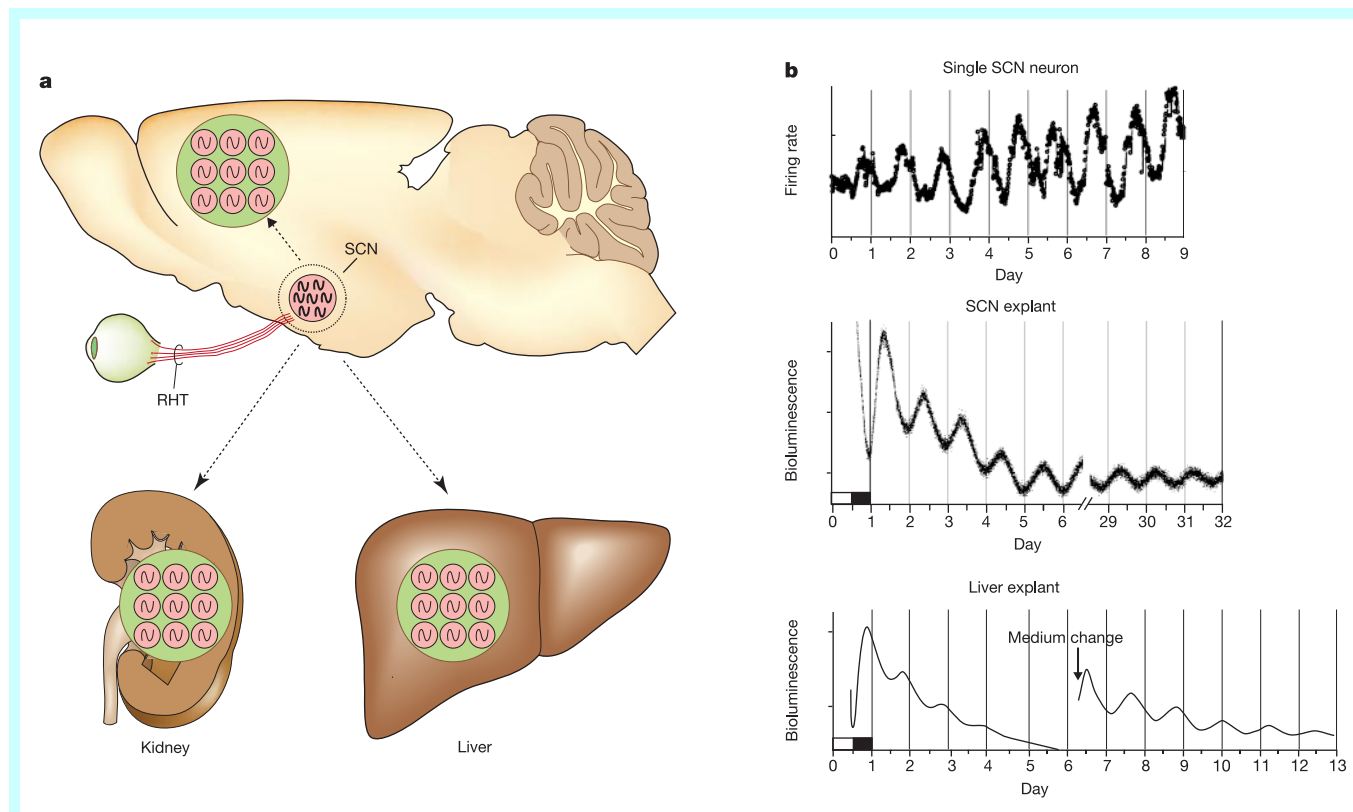


Figure 1 The mammalian circadian timing system is a hierarchy of dispersed oscillators. **a**, The master clock in the SCN is composed of numerous clock cells. The SCN receives light information by a direct retinohypothalamic tract (RHT) to entrain the clock to the 24-h day. The entrained SCN, in turn, coordinates the timing of slave oscillators in other brain areas (for example, cortex) and in peripheral organs (for example, kidney and liver). **b**, A single SCN neuron in culture expresses robust circadian rhythms in firing rate over 9 days of study, proving that the core clock

mechanism is contained within single cells (adapted from ref. 83). SCN and liver explants from transgenic rats expressing a *mPer1*-driven luciferase reporter gene exhibit bioluminescence rhythms in culture; the black and white bars along the x axis indicate the light–dark cycle at the time of tissue collection (adapted from ref. 9). The SCN explant rhythm persists for weeks in culture, whereas the liver explant rhythm dampens. A medium change on day 7 restarts the liver oscillation, showing that the dampening was not due to tissue death.

genetic basis of circadian behaviour may help to decipher this connection for more complex behaviours. Understanding the molecular clock could increase our knowledge of how gene mutations of the molecular clock contribute to psychopathology (for example, major depression and seasonal affective disorder)². Similarly, such understanding should lead to new strategies for pharmacological manipulation of the human clock to improve the treatment of jet lag and ailments affecting shift workers, and of clock-related sleep and psychiatric disorders.

A hierarchy of distributed oscillators

Circadian timing in mammals is organized in a hierarchy of multiple circadian oscillators (Fig. 1a). The oscillatory machinery of the master clock is contained within single neurons³ (Fig. 1b), and it is possible that most of the approximately 20,000 neurons that comprise the bilateral SCN are 'clock cells'. Molecular evidence is beginning to emerge for functionally distinct populations of clock cells within the SCN⁴⁻⁶.

Intriguingly, there are also circadian oscillators scattered throughout the body, as the genes involved in the intracellular SCN clock mechanism have been found to be rhythmically expressed in other brain areas, in peripheral organs, and even in immortalized cell lines in culture⁷⁻⁹. The extra-SCN, 'slave' oscillators expressed *in vivo* can only sustain 24-h oscillations for a few days without input from the master clock (Fig. 1b). For orchestrated circadian timing, the collective SCN synchronizes the timing of slave oscillators, each of which is a multioscillatory entity. Synchronized slave oscillators, in turn, regulate local rhythms in physiology and behaviour. A hierarchical multioscillatory system seems to confer precise phase control and stability on the widely distributed physiological systems it regulates¹⁰.

Analysis of clockwork function in slave oscillators (such as the liver and cultured cell lines) has been invaluable for defining biochemical principles important for the core clock mechanism, as rigorous biochemical analysis of a small brain structure like the SCN is difficult. Indeed, genetic studies have shown that the molecular composition of the timing mechanisms in the SCN clock and slave oscillators is very similar¹¹. The actual mechanism that distinguishes the self-sustaining oscillatory function of the master clock from the damped oscillation of slave oscillators is unknown, but the mechanism may have more to do with global differences in clock protein levels and/or kinetics than the existence of a specific element (gene/protein) expressed only in the SCN.

Transcriptional feedback loops

Knowledge of circadian clock mechanisms in the fruit fly *Drosophila melanogaster* has greatly aided the formulation of mammalian clock mechanisms¹². Homologues of most of the genes involved in the fly

circadian clock have been cloned in mammals, and the general core clock mechanism of interacting transcriptional feedback loops is similar between flies and mice. However, there has been a shuffling of specific functions between several structurally disparate components, and gene duplication has led to increased complexity among mammalian clock genes^{1,12}. Even though the principal clock genes may have been identified (Table 1), genome-wide complex trait analysis in mice has revealed the presence of yet-to-be discovered clock-modulating genes¹³.

The intracellular clock mechanism in the mouse involves interacting positive and negative transcriptional feedback loops that drive recurrent rhythms in the RNA and protein levels of key clock components (Fig. 2). Rhythmic transcriptional enhancement by two basic helix-loop-helix (bHLH)-PAS (Period-Arnt-Single-minded)-containing transcription factors, CLOCK and BMAL1 (also called MOP3), is essential for clockwork function and provides the basic drive to the system¹⁴⁻¹⁷. CLOCK-BMAL1 heterodimers activate transcription by binding to E box enhancers and are highly selective for those with the nucleotide sequence CACGTG^{15,16}. Negative feedback involves rhythmic inhibition of the CLOCK-BMAL1 drive by negative regulators (Fig. 2). Specifically, CLOCK-BMAL1 heterodimers activate the rhythmic transcription of three *period* genes (*mPer1*–*mPer3* in the mouse) and two *cryptochrome* genes (*mCry1* and *mCry2*). The resultant mPER and mCRY proteins translocate back into the nucleus where the mCRY proteins act as negative regulators by directly interacting with CLOCK and/or BMAL1 to inhibit transcription, closing the negative feedback loop¹⁸⁻²¹.

The positive feedback loop involves the rhythmic regulation of *Bmal1* transcription, whose RNA levels peak 12 h out of phase relative to *mPer* and *mCry* RNAs^{21,22}. Recent studies provide a revised view of the regulation of the positive feedback loop (Fig. 2): to generate the positive loop, CLOCK-BMAL1 heterodimers, while activating *mPer* and *mCry* transcription, also activate transcription of the orphan nuclear receptor gene *Rev-Erbα*²³. The REV-ERBα protein then represses *Bmal1* transcription by acting through Rev-Erb/ROR response elements in its promoter^{23,24}. As a result, *Bmal1* RNA levels fall, whereas *mPer* and *mCry* RNA levels rise. When the mCRY proteins enter the nucleus to inhibit *mPer* and *mCry* transcription (through actions on CLOCK-BMAL1), they also inhibit *Rev-Erbα* transcription resulting in a de-repression (activation) of *Bmal1* transcription^{23,25}. mPER2 may provide the positive drive on *Bmal1* transcription when *Rev-Erbα* expression is inhibited²¹. Thus, the positive and negative transcriptional feedback loops are co-regulated by CLOCK-BMAL1 heterodimers, with differences in the loop dynamics coming from the differential activities of the co-regulated protein products (Fig. 2).

mPER and mCRY outputs of the negative feedback loop are

Table 1 Components of the mouse clock mechanism

Proteins		Mutations*	
Family	Member	Gene(s)	Behavioural phenotype†
bHLH-PAS	CLOCK‡	<i>Clock</i> §	Long period, then arrhythmic
	BMAL1 (MOP3)	<i>Bmal1</i> (Mop3)	Arrhythmic
PER-PAS	PER1	<i>Per1</i> or <i>Per1</i> + <i>Per3</i>	Short period, then arrhythmic
	PER2	<i>Per2</i> or <i>Per2</i> + <i>Per3</i>	Short period, then arrhythmic
	PER3	<i>Per3</i>	Short period
Flavoproteins	CRY1	<i>Per1</i> + <i>Per2</i>	Arrhythmic
	CRY2	<i>Cry1</i>	Short period
		<i>Cry2</i>	Long period
Casein kinase	CKIε	<i>Cry1</i> + <i>Cry2</i>	Arrhythmic
Orphan nuclear receptor	REV-ERBα	<i>CKIε</i>	Short period
		<i>Rev-Erbα</i>	Short period

* Unless otherwise noted, the mutations listed are deletion mutations induced by targeted mutagenesis.

† The most severe phenotypes of homozygous mutant animals studied under constant conditions are listed.

‡ A CLOCK-like role has been described for MOP4 (also known as NPAS2)—a bHLH-PAS transcription factor closely related to CLOCK—in the cerebral cortex³² and in the vasculature⁴¹.

§ Ethylnitrosourea-induced semidominant autosomal mutation. The mutation is a nucleotide transversion in a splice donor site causing exon skipping and deletion of part of the transactivation domain¹⁴.

|| Spontaneous semidominant autosomal mutation described in the Syrian hamster (the *tau* mutation). The mutant enzyme is deficient in its ability to phosphorylate PER³⁵.

essential for maintaining a functioning circadian clock, as disruption of either the *mPer1* and *mPer2* genes together or both *mCry* genes causes immediate behavioural arrhythmicity when the double-knockout animals are placed in constant conditions^{19,26–28} (Table 1). There is partial compensation of function between family members, as the clock continues to oscillate after single gene mutations, albeit for variable periods of time in constant conditions depending on the gene mutated^{19,26–30} (Table 1). mPER3 does not have a critical role in the maintenance of the core clock feedback loops, as molecular and behavioural rhythms are preserved in *mPer3* null mutant mice, and there is no synergy with either the homozygous *mPer1* or *mPer2* mutants^{26,31} (Table 1). Instead, mPER3 probably functions as an output signal.

Post-translational control of the clockwork

Circadian patterns of clock protein abundance, phosphorylation, interactions and subcellular location each contribute to building time delays into the 24-h clockwork. Although translational control may also contribute to a time delay between *mPer* RNA and protein rhythms, post-translational processes have been the most well studied mechanisms for post-transcriptional control.

Clock protein phosphorylation

CLOCK, BMAL1, mPER1 and mPER2 undergo temporal changes in phosphorylation *in vivo*, with maximal phosphorylation correlating with the time of negative feedback on *mPer/mCry* transcription³² (Figs 2 and 3). Ironically, transcriptional activation occurs when levels of CLOCK and BMAL1 in the nucleus are at their

lowest³², suggesting that the phosphorylation status of the transcription factors is important for the transcriptional competency of the heterodimer^{33,34}. Phosphorylation may also be important for the translocation of CLOCK and BMAL1 into the nucleus, and the formation of protein complexes that inhibit CLOCK–BMAL1-mediated transcription (see below).

So which kinases are important for the mammalian clockwork? Casein kinase I ϵ (CKI ϵ) is an important clock modulator in Syrian hamsters, because a defect in the hamster CKI ϵ gene corresponds to the short-period *tau* mutation³⁵. CKI ϵ can phosphorylate PER1 and PER2, and the mutant hamster kinase has a lower rate of phosphorylation³⁵. In addition, a human genetic disorder characterized by shortened circadian period and advanced sleep phase is associated with a missense mutation in human PER2, and the mutant protein is less effectively phosphorylated by CKI ϵ *in vitro*³⁶. Of note, the degree of PER1 and PER2 phosphorylation is not appreciably altered in homozygous *tau* mutant hamsters³². With the recent finding that CKI ϵ can also phosphorylate mCRY1, mCRY2 and BMAL1, it is possible that alterations in the temporal phosphorylation of one of these proteins may contribute to the altered clock phenotype in *tau* mutant hamsters³³.

In vivo evidence suggests that CKI δ is a second kinase important for the mammalian clockwork³². CKI δ is highly homologous to CKI ϵ (76% identical at the amino acid level in humans) and phosphorylates mPER1, mPER2, mCRY1, mCRY2 and BMAL1 *in vitro*^{33,37}. Mitogen-activated protein kinase can also phosphorylate BMAL1 *in vitro* (ref. 34). Recent work in *Drosophila* suggests that glycogen synthase kinase-3 and related kinases should be examined for their involvement in the mammalian clockwork³⁸. The kinases responsible for rhythmic phosphorylation of CLOCK have not been identified. Clearly, the orchestrated temporal programme of clock protein phosphorylation contributes to the 24-h time kinetic of the clockwork, requires the coordinated activity of several kinases, and probably involves phosphatases.

Regulated nuclear entry

Nuclear entry of the mPER and mCRY proteins is a vital checkpoint for progression of the clockwork cycle. *In vivo* studies of the liver oscillator show that the mPER proteins are rate limiting for the mPER–mCRY interactions in cytoplasm that, in turn, are necessary for nuclear accumulation of the complex³². Accordingly, the robust oscillation in mPER protein abundance drives the clock mechanism forward, as it brings clock protein complexes into the nucleus at the proper time for negative transcriptional feedback. There is also a co-dependency between the mPER and mCRY proteins for effective nuclear accumulation. This helps explain why the molecular clock continues to cycle with single gene mutations (*mPer1*, *mPer2*, *mCry1* or *mCry2*) but abruptly stops cycling in double-knockout animals in which either *mPer1* plus *mPer2* or *mCry1* plus *mCry2* genes are targeted (Table 1). The balance between the nuclear import and export of clock proteins also contributes to their cellular location and may provide a point for fine-tuning circadian cycle length^{39,40}.

Rhythms in subcellular localization of CKI ϵ and CKI δ are synchronous and are driven by association with mPER and mCRY occurring in the cytoplasm, which enable subsequent accumulation of multimeric complexes in the nucleus³² (Fig. 2). PER proteins, which have distinct binding sites for mCRY and CKI ϵ /CKI δ , act as bridge proteins that are not only required for the trimeric protein assembly, but may also be required for phosphorylation of the mCRY proteins by the casein kinases^{33,37}.

Of note, the mCRY proteins are essential for the stability of phosphorylated mPER2, probably through direct mPER2–mCRY association³². This interaction seems to protect phosphorylated mPER2 from ubiquitination and subsequent degradation in the proteasome⁴⁰. On the other hand, the mCRY proteins are not necessary for the stabilization of phosphorylated mPER1 (refs 21,

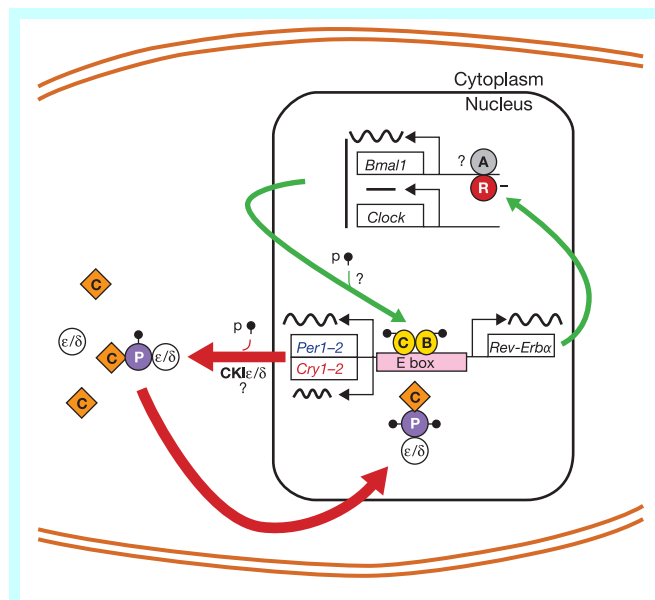


Figure 2 Mammalian circadian clockwork model. The clock mechanism comprises interactive positive (green) and negative (red) feedback loops. CLOCK (C, oval) and BMAL1 (B, oval) form heterodimers and activate transcription of the *Per*, *Cry* and *Rev-Erbα* genes through E-box enhancers. As the levels of PER proteins increase (P, blue circle), they complex with CRY proteins (C, diamond) and CKI ϵ /CKI δ (ε/δ, circle), and are phosphorylated (p). In the nucleus, the CRY–PER–CKI ϵ /CKI δ complexes associate with CLOCK–BMAL1 heterodimers to shut down transcription while the heterodimer remains bound to DNA, forming the negative feedback loop. For the positive feedback loop, increasing REV-ERBα levels (R, circle) act through Rev-Erb/ROR response elements in the *Bmal1* promoter to repress (–) *Bmal1* transcription. CRY-mediated inhibition of CLOCK–BMAL1-mediated transcription de-represses (activates) *Bmal1* transcription, because REV-ERBα-mediated repression is inhibited. An activator (A, circle) may positively regulate *Bmal1* transcription (?) alone or by interacting with mPER2. There are probably kinases (?) other than CKI ϵ and CKI δ that participate in phosphorylation of clock proteins.

32). It is not known how the differential effects of mCRY protection contribute to the roles of mPER1 and mPER2 in the clock mechanism.

A nuclear timesome

Once in the nucleus, mCRY–mPER–CKI ϵ /CK1 δ complexes associate with CLOCK and BMAL1 (ref. 32). This ‘timesome’ negatively regulates transcription of *mPer*, *mCry* and *Rev-Erb α* genes through disruption of activity of the transcriptional complex (Fig. 3). Chromatin immunoprecipitation experiments show that CLOCK and BMAL1 are both constitutively bound to *mPer1* E boxes over the circadian cycle in the liver oscillator³². It is the rhythmic binding of negative regulators to the DNA-anchored CLOCK–BMAL1 heterodimers that generates the rhythm in transcriptional activity³². In slave oscillators in vascular tissue, the activation of the nuclear hormone receptors retinoic acid receptor- α (RAR α) and retinoid-X receptor- α (RXR α) seems to act cooperatively with the mCRY proteins to negatively regulate CLOCK–BMAL1-mediated transcription⁴¹. This result reveals potential complexity provided by the involvement of numerous co-regulators. The constant binding of CLOCK and BMAL1 to DNA argues against a major role of redox state in modifying the binding of CLOCK–BMAL1 heterodimers to E boxes *in vivo*^{42,43}.

The hyperphosphorylated state of the timesome probably targets the clock protein complex for degradation at the end of the inhibitory phase, perhaps through the ubiquitin-proteasome pathway^{37,40}. The continued presence of CLOCK–BMAL1 heterodimers bound to DNA at the end of the negative phase of transcriptional regulation probably reflects a mixed population of newly synthesized and ‘older’ transcriptional complexes, which differ in functional activity.

Photic input to the SCN

A ‘circadian’ photoreceptor

A critical feature of circadian timing is the ability of the clockwork to be reset by environmental stimuli. In mammals, light is the most potent entraining signal, with the retinohypothalamic tract (RHT) being the principal retinal pathway through which entraining information reaches the SCN^{44,45}. Recent studies build a strong

case for the involvement of the photopigment melanopsin in circadian photoreception.

Classical retinal photoreceptors, the rods and cones, with their opsin-based visual pigments are necessary for the conscious perception of light; however, they are dispensable for several light responses⁴⁵. Light-induced phase shifts of the SCN clock, inhibition of nocturnal melatonin production, inhibition of activity by light (negative masking), and pupillary constrictor reflexes are all mediated by a non-rod, non-cone system. These responses have similar properties, including high threshold and the ability to integrate photic input over substantial periods⁴⁵. Isolation of novel opsin-like molecules fuelled speculation that a new photoreceptive molecule present within the inner retina mediates these responses.

Localization of melanopsin gene expression and melanopsin immunoreactivity within a widely dispersed population of retinal ganglion cells suggests that melanopsin is a circadian photoreceptor^{46–48}. This special population of retinal ganglion cells is unique in neurochemical phenotype (containing pituitary adenylyl cyclase-activating peptide (PACAP) as well as melanopsin), has the appropriate morphology (with large dendritic fields), and projects directly to the SCN and intergeniculate leaflet^{47,49,50}. Most importantly, melanopsin-positive ganglion cells projecting to the SCN respond directly to light, even when physically isolated^{47,51}. Although these studies do not exclude that another molecule within the melanopsin-positive cells is responsible for photosensitivity, the most parsimonious view is that the non-rod, non-cone system conveying light to the SCN is mediated by melanopsin action within this special subset of ganglion cells (Fig. 4).

However, it is not clear whether melanopsin is sufficient for circadian photoreception. The melanopsin pathway to the SCN may be complemented by the classical photoreceptor system and/or retinal cryptochromes^{52,53}, resulting in functional redundancy. Even though the mCRY proteins bind flavin, a light-sensing role has not been shown⁵⁴. Rather, as detailed previously, the primary function of the mCRY proteins is in negative regulation within the clock feedback loops.

Transducing retinal input

SCN neurons downstream of retinal ganglion cell terminals receive and process photic input. The principal neurotransmitters of the retinohypothalamic tract are glutamate and PACAP. Photic and non-photic input also reaches the SCN indirectly through the intergeniculate leaflet and midbrain, with GABA (γ -aminobutyric acid), neuropeptide Y and serotonin having principal roles. The neurochemistry of input to the SCN has been reviewed elsewhere¹.

How does activation of retinal ganglion cells lead to alterations in the molecular clock in SCN neurons? Gene expression of *mPer1* is rapidly induced after exposure to light at either the beginning or the end of night⁵⁵. Whereas *mPer2* induction is robust after exposure to light early in the night, exposure later in the night does not lead to a readily detectable induction⁸. These data focus attention on *mPer2* as a mediator of early-night responses (phase delays) and on *mPer1* as a mediator of phase advances. One study of mice with targeted disruption of either *mPer1* or *mPer2* is consistent with these proposed roles⁵⁶. Another study using a similar strategy, however, shows that the *mPer1* gene is not necessary for light-induced phase shifts³⁰. The precise roles of the *mPer* genes in light-induced clock resetting need clarification.

Analysis of clock proteins after nocturnal light exposure indicates that the number of nuclei immunoreactive for mPER1 and mCRY1 is increased in the SCN several hours after a phase-resetting light pulse⁵⁷. The finding that mCRY1 is increased along with mPER1, despite the absence of an effect of light on *mCry1* RNA levels, suggests that more mCRY is translocated to the nucleus owing to increasing mPER levels. Thus, PER-dependent nuclear accumulation of CRY may be key for resetting circadian phase, as well as

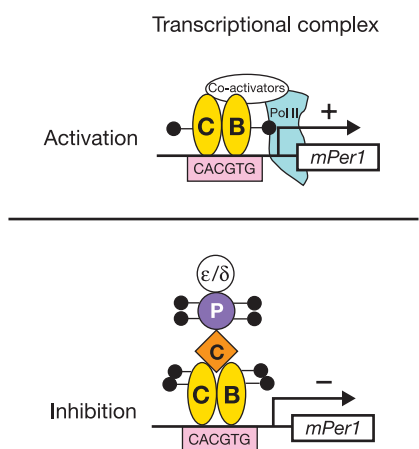


Figure 3 CLOCK–BMAL1 heterodimers remain bound to E boxes over the circadian cycle. During transcriptional activation, CLOCK (C) and BMAL1 (B) are bound to CACGTG E boxes in the *mPer1* promoter, and co-activators are recruited to activate transcription. During transcriptional inhibition, the CRY-containing complex of negative regulators binds to the CLOCK–BMAL1 heterodimer, and may inactivate the co-activator complex. At the time of transcriptional inhibition, CLOCK, BMAL1 and the mPER proteins are hyperphosphorylated.

controlling the central clock mechanism^{32,40}.

The mechanisms by which light leads to an increase in *mPer1* gene expression are better understood. Photic induction of *mPer1* and *mPer2* gene expression seems to be mediated by chromatin remodelling⁵⁸ and the binding of phosphorylated CREB (cyclic AMP responsive element-binding protein; pCREB) to a cAMP-responsive element (CRE) in the respective promoters⁵⁹ (Fig. 4b). Notably, the photic induction is largely independent of E-box-mediated enhancement, as light induces *Per* expression quite effectively in *Clock/Clock* mutant mice⁶⁰. Conversely, activation of *mPer1* and *mPer2* gene expression by CLOCK–BMAL1 heterodimers (presumably acting on E boxes) is independent of the CRE elements present within these promoters⁵⁹. One unique feature of the circadian response is that CREB phosphorylated on serine 142 (rather than the more widely studied Ser 133) seems to be primarily responsible for photic induction of *mPer1* RNA⁶¹.

SCN to slave oscillators to local rhythms

Sodium-dependent action potentials provide the primary means by which the SCN transmit circadian outputs to other brain areas and are essential to its role as a pacemaker¹. There are at least two ionic mechanisms that are under clock control in the SCN—an L-type Ca^{2+} current and a K^{+} current that is essential for maintaining membrane potential⁶². The summed activity of these currents may contribute to the striking firing rate rhythm exhibited by SCN neurons³. But how does the core clock mechanism regulate these ionic events? Channel subunits could be regulated rhythmically at the transcriptional or translational level. Channel activity could also be regulated by a rhythm in post-translational processes, such as phosphorylation and/or interactions with regulatory proteins. These regulator proteins could be clock-controlled genes (CCGs) that contain E-box enhancers and are therefore directly controlled by CLOCK–BMAL1 heterodimers (so-called first-order CCGs), or downstream CCGs indirectly controlled by CLOCK–BMAL1 heterodimers.

Signalling molecules from SCN efferents include neurotransmitters and secreted factors. An output role for secreted factors comes from the results of SCN transplant studies that suggest that the alternating activity of SCN-derived ‘inhibitory’ and ‘activating’ factors drives locomotor activity rhythms (and hence rest/activity and sleep/wake cycles) in rodents. Two neuropeptides, transforming growth factor- α (TGF- α) and prokineticin-2 (PK2), have recently been identified as candidate output factors. TGF- α is a potential inhibitory substance, as its infusion into the third ventricle inhibits locomotor activity⁶³. The growth factor is expressed in the SCN and retina and appears to inhibit locomotor activity by action on receptors in the hypothalamic subparaventricular zone (SPZ), a major relay station for SCN efferents. The most marked effect of growth factor signalling, however, is regulating light-induced suppression of locomotion, independent of circadian control⁶³. By mediating this light-induced suppression (so-called ‘masking’), TGF- α reinforces circadian control of the nocturnal activity profile.

PK2, a peptide whose RNA levels are highly and rhythmically expressed in mouse SCN, contributes more directly to the circadian control of behavioural activity levels⁶⁴. PK2 is a first-order CCG in the SCN, as its RNA levels are regulated by CLOCK and BMAL1 acting on E-box enhancers in the gene’s promoter. Receptors for PK2 are expressed in many SCN target sites, as well as in the SCN itself, but notably not in the SPZ. Intracerebroventricular infusion of PK2 at night, when peptide levels are normally low, markedly reduces the nocturnal increase in locomotion⁶⁴. The PK2-induced decrease in locomotor activity is consistent with a role of the endogenous peptide in suppressing locomotion during the day, when mice are normally inactive. Thus, clock-controlled PK2 expression sculpts the daily activity profile and may provide a key link between the circadian clock and behavioural outputs (including sleep).

What are the mechanisms whereby the SCN pacemaker controls slave oscillators in tissues outside of the brain? There are neural outputs from the SCN to peripheral organs by means of the

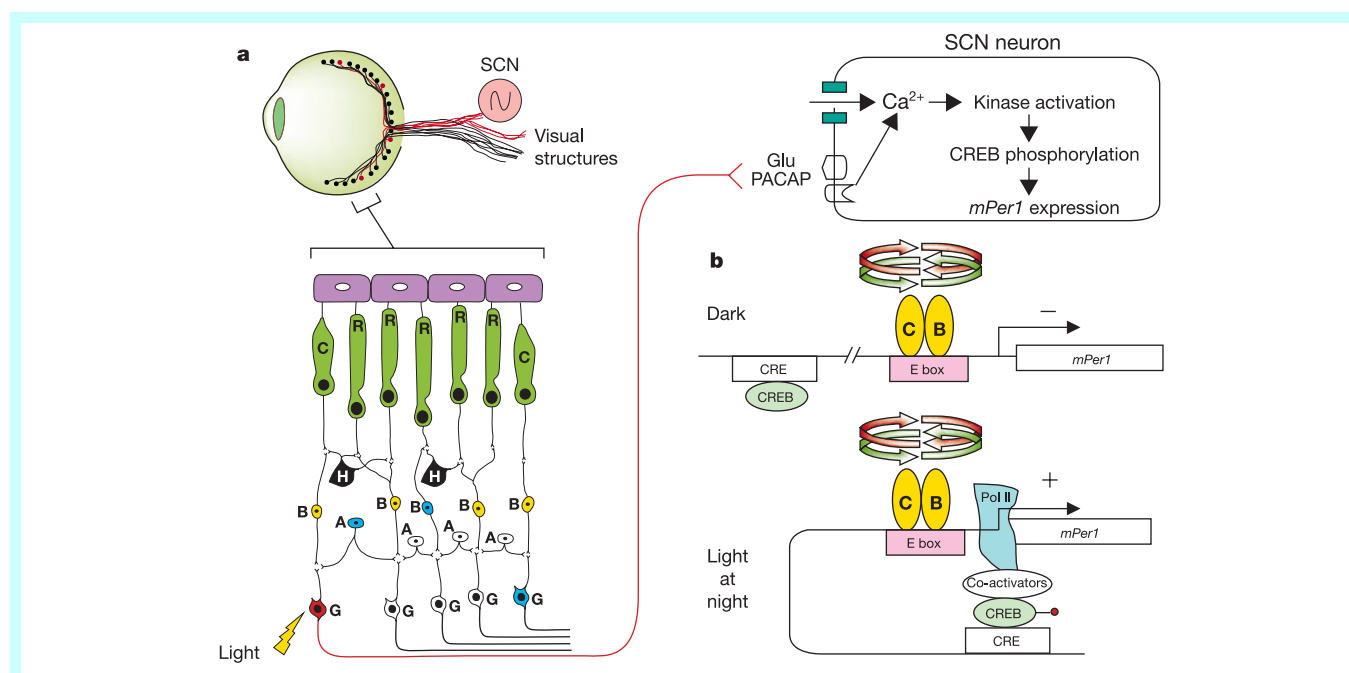


Figure 4 New visual pathway from retina to the SCN. **a**, A small population of widely dispersed, melanopsin-positive ganglion cells (red) form the retinohypothalamic tract that projects to the SCN. These ganglion cells (G) are directly light responsive. They also receive input from rods (R) and cones (C) through bipolar (B) and amacrine cells (A), some of which may contain cryptochromes (blue). The precise anatomy of inputs to the

melanopsin-positive neurons remains to be established. Glutamate (Glu) and PACAP mediate light effects on the *mPer* genes in SCN neurons. **b**, Light at night activates *mPer1* expression through phosphorylated CREB, independent of CLOCK–BMAL1 E-box control. Red and green arrows indicate interacting negative and positive loops of the clockwork, respectively.

autonomic nervous system, supporting direct (multisynaptic) neural control⁶⁵. Hormonal signals are capable of entraining peripheral oscillators, as glucocorticoid agonists can effectively shift peripheral oscillators in mice⁶⁶. There is also a more indirect way by which the SCN control the phase of slave oscillators in peripheral tissues. The SCN entrain some slave oscillators (for example, liver) by regulating the rest/activity cycle and thereby the timing of feeding behaviour^{67,68}. Thus an artificially imposed feeding schedule in rodents can completely uncouple peripheral oscillators from SCN control. Feeding-induced signals, antagonized by glucocorticoids, entrain these peripheral oscillators⁶⁹. The complex interaction of neural, hormonal and behavioural outputs may regulate the timing of slave oscillators by converging on *Per1* and *Per2* gene expression in peripheral tissues (Fig. 5a).

Once the timing of slave oscillators is coordinated by the SCN, there needs to be a mechanism to transduce the synchronized molecular oscillation into local rhythms. One way this could happen is through local first-order CCGs^{70,71} (Fig. 5b). D-element binding protein (DBP) is the prototype of such a CCG, as it regulates the rhythmic transcription of key enzymes involved in hepatic metabolic processes⁷². Recent evidence indicates that DBP works together with the related basic leucine zipper transcription factor E4BP4, whose rhythm phase is opposite that of DBP, to affect differentially the same *cis*-acting element and thereby drive rhythmicity in responsive genes⁷³. E4BP4 rhythmicity is probably regulated by the repressor REV-ERB α , similar to the way *Bmal1* rhythmicity is regulated in the core clock mechanism²⁴.

The combination of DNA microarray technology and the sequencing of the genomes of humans and rodents provides a powerful way

to identify the multitude of CCGs in mammalian tissues⁷⁴. Microarray analysis in fact has been applied to the study of circadian gene expression in serum-shocked fibroblast cultures^{75,76} and in various tissues in rats and mice^{24,77–80}. Collectively, these studies show that between 2 and 10% of the analysed genes exhibit circadian oscillations in steady-state RNA levels. Circadian gene expression is remarkably tissue-specific (less than 5% overlap between tissues) and, in many cases, involves rate-limiting steps that are distinct for the functions of that organ^{78,79}. Unexpectedly, only a small subset of rhythmic genes seems to be under direct transcriptional control by CLOCK–BMAL1 heterodimers⁷⁹. These first-order CCGs (such as PK2 and DBP), along with the core clock proteins, thus seem to be critical mediators of circadian information, as they are the most direct way through which the core oscillation can be transduced to regulate downstream events. The microarray data provide direct evidence for the ubiquitous role that circadian timing has in regulating diverse physiological events in mammals.

What's next?

Although we can assemble the cloned clock genes and their protein products into a coherent clockwork model, the model continues to evolve (for example, a revised view of the regulation of the positive feedback loop; see Fig. 2). Moreover, it is likely that large parts of the mammalian clockwork remain to be unravelled. A case in point is the unexpected loss of robust behavioural and molecular rhythmicity in mice with targeted disruption of the VPAC₂ receptor⁸¹. Intercellular signalling apparently has an important paracrine role in reinforcing intracellular circadian oscillations. Understanding the mechanisms by which single oscillators interact to form a functional oscillator at the tissue level remains one of the important challenges for future studies.

Recent DNA microarray data suggest that direct regulation of rhythmic gene expression by the core feedback loop is a rare, but critical, event controlling the distinctive temporal profiles of gene expression in peripheral tissues. The use of chromatin immunoprecipitation combined with DNA microarray analysis of the regulatory regions of mammalian genes provides a focused approach for discovery of first-order CCGs directly under CLOCK–BMAL1 control in the SCN and in slave oscillators. Fertile research also lies in elucidating the cascading interactions of networks of CCGs that connect the clockwork to the expressed rhythms. Finally, work is needed to assess the importance of locally controlled, rhythmic gene expression in physiology and how loss of circadian control contributes to disease states at the organ and systemic levels. □

doi:10.1038/nature00965.

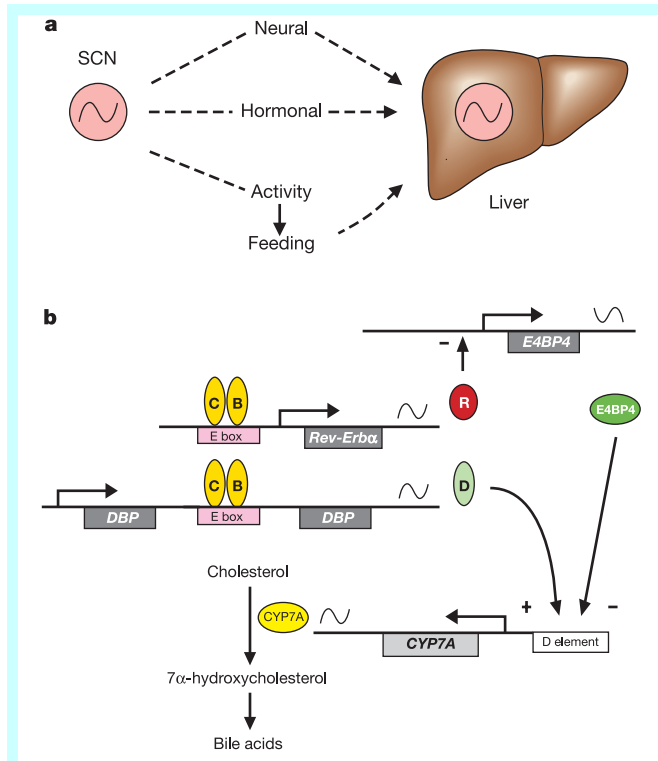


Figure 5 Mechanisms from SCN to clock-controlled gene expression in liver. **a**, Multiple output mechanisms from SCN to liver. Adapted from ref. 69. **b**, DBP-regulated local output gene in liver. DBP (D) rhythmicity is driven directly by CLOCK–BMAL1 heterodimers through an intronic E-box enhancer. E4BP4 rhythmicity is antiphase and driven by the repressor REV-ERB α (R). DBP (+) and E4BP4 (–) act through a D element to coordinate the rhythmic transcription of cholesterol 7 α -hydroxylase (*CYP7A*), which is the rate-limiting step for bile acid production.

- Reppert, S. M. & Weaver, D. R. Molecular analysis of mammalian circadian rhythms. *Ann. Rev. Physiol.* **63**, 647–676 (2001).
- Bunney, W. E. & Bunney, B. G. Molecular clock genes in man and lower animals: possible implications for circadian abnormalities in depression. *Neuropsychopharmacology* **22**, 335–345 (2000).
- Welsh, D. K., Logothetis, D. E., Meister, M. & Reppert, S. M. Individual neurons dissociated from rat suprachiasmatic nucleus express independently phase circadian firing rhythms. *Neuron* **14**, 697–706 (1995).
- Jagota, A., de la Iglesia, H. O. & Schwartz, W. J. Morning and evening circadian oscillations in the suprachiasmatic nucleus *in vitro*. *Nature Neurosci.* **3**, 372–376 (2000).
- Low-Zeddis, S. S. & Takahashi, J. S. Chimera analysis of the Clock mutation in mice shows that complex cellular integration determines circadian behaviour. *Cell* **105**, 25–42 (2001).
- Hamada, T., LeSauter, J., Venuti, J. M. & Silver, R. Expression of Period genes: rhythmic and nonrhythmic compartments of the suprachiasmatic nucleus pacemaker. *J. Neurosci.* **21**, 7742–7750 (2001).
- Balsalobre, A., Damiola, F. & Schibler, U. A serum shock induces circadian gene expression in mammalian tissue culture cells. *Cell* **93**, 929–937 (1998).
- Zylka, M. J., Shearman, L. P., Weaver, D. R. & Reppert, S. M. Three *period* homologs in mammals: differential light responses in the suprachiasmatic circadian clock and oscillating transcripts outside of brain. *Neuron* **20**, 1103–1110 (1998).
- Yamazaki, S. *et al.* Resetting central and peripheral circadian oscillators in transgenic rats. *Science* **288**, 682–685 (2000).
- Pando, M. P., Morse, D., Cermakian, N. & Sassone-Corsi, P. Phenotypic rescue of a peripheral clock genetic defect via SCN hierarchical dominance. *Cell* **110**, 107–117 (2002).
- Yagita, K., Tamanini, F., van Der Horst, G. T. & Okamura, H. Molecular mechanisms of the biological clock in cultured fibroblasts. *Science* **292**, 278–281 (2001).

12. Young, M. W. & Kay, S. A. Time zones: a comparative genetics of circadian clocks. *Nature Rev. Genet.* **2**, 702–715 (2001).
13. Shimomura, K. *et al.* Genome-wide epistatic interaction analysis reveals complex genetic determinants of circadian behaviour in mice. *Genome Res.* **11**, 959–980 (2001).
14. King, D. P. *et al.* Positional cloning of the mouse circadian *Clock* gene. *Cell* **89**, 641–653 (1997).
15. Gekakis, N. *et al.* Role of the *CLOCK* protein in the mammalian circadian mechanism. *Science* **280**, 1564–1569 (1998).
16. Hogenesch, J. B., Gu, Y.-Z., Jain, S. & Bradfield, C. A. The basic helix-loop-helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors. *Proc. Natl Acad. Sci. USA* **95**, 5474–5479 (1998).
17. Bunger, M. K. *et al.* *Mop3* is an essential component of the master circadian pacemaker in mammals. *Cell* **103**, 1009–1017 (2000).
18. Kume, K. *et al.* mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop. *Cell* **98**, 193–205 (1999).
19. Vitaterna, M. H. *et al.* Differential regulation of mammalian *period* genes and circadian rhythmicity by *cryptochromes* 1 and 2. *Proc. Natl Acad. Sci. USA* **96**, 12114–12119 (1999).
20. Okamura, H. *et al.* Photic induction of mPer1 and mPer2 in cry-deficient mice lacking a biological clock. *Science* **286**, 2531–2534 (1999).
21. Shearman, L. P. *et al.* Interacting molecular loops in the mammalian circadian clock. *Science* **288**, 1013–1019 (2000).
22. Oishi, K., Fukui, H. & Ishida, N. Rhythmic expression of BMAL1 mRNA is altered in *Clock* mutant mice: differential regulation in the suprachiasmatic nucleus and peripheral tissues. *Biochem. Biophys. Res. Commun.* **268**, 164–171 (2000).
23. Preitner, N. *et al.* The orphan nuclear receptor REV-ERBA controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell* **110**, 251–260 (2002).
24. Ueda, H. R. *et al.* A transcription factor response element for gene expression during circadian night. *Nature* **418**, 534–539 (2002).
25. Yu, W., Nomura, M. & Ike, M. Interacting feedback loops within the mammalian clock: BMAL1 is negatively autoregulated and upregulated by CRY1, CRY2, and PER2. *Biochem. Biophys. Res. Commun.* **290**, 933–942 (2002).
26. Bae, K. *et al.* Differential functions of mPer1, mPer2, and mPer3 in the SCN circadian clock. *Neuron* **30**, 525–536 (2001).
27. Zheng, B. *et al.* Nonredundant roles of the mPer1 and mPer2 genes in the mammalian circadian clock. *Cell* **105**, 683–694 (2001).
28. van der Horst, G. T. J. *et al.* Mammalian *Cry1* and *Cry2* are essential for maintenance of circadian rhythms. *Nature* **398**, 627–630 (1999).
29. Zheng, B. *et al.* The mPer2 gene encodes a function component of the mammalian clock. *Nature* **400**, 169–173 (1999).
30. Cermakian, N., Monaco, L., Pando, M. P., Dierich, A. & Sassone-Corsi, P. Altered behavioural rhythms and clock gene expression in mice with a targeted mutation in the *Period1* gene. *EMBO J.* **20**, 3967–3974 (2001).
31. Shearman, L. P., Jin, X., Lee, C., Reppert, S. M. & Weaver, D. R. Targeted disruption of the mPer3 gene: Subtle effects on circadian clock function. *Mol. Cell. Biol.* **20**, 6269–6275 (2000).
32. Lee, C., Etchegaray, J. P., Cagampang, F. R., Loudon, A. S. & Reppert, S. M. Posttranslational mechanisms regulate the mammalian circadian clock. *Cell* **107**, 855–867 (2001).
33. Eide, E. J., Vielhaber, E. L., Hinz, W. A. & Virshup, D. M. The circadian regulatory protein BMAL1 and cryptochromes are substrates of casein kinase 1 ϵ (CK1 ϵ). *J. Biol. Chem.* **277**, 17248–17254 (2002).
34. Sanada, K., Okano, T. & Fukada, Y. Mitogen-activated protein kinase phosphorylates and negatively regulates basic helix-loop-helix-PAS transcription factor BMAL1. *J. Biol. Chem.* **277**, 267–271 (2002).
35. Lowrey, P. L. *et al.* Positional syntenic cloning and functional characterization of a mammalian circadian mutation *tau*. *Science* **288**, 483–491 (2000).
36. Toh, K. L. *et al.* An *hPer2* phosphorylation site mutation in familial advanced sleep phase syndrome. *Science* **291**, 1040–1043 (2001).
37. Akashi, M., Tsuchiya, Y., Yoshino, T. & Nishida, E. Control of intracellular dynamics of mammalian period proteins by casein kinase I epsilon (CK1 ϵ) and CK1 δ in cultured cells. *Mol. Cell. Biol.* **22**, 1693–1703 (2002).
38. Martinek, S., Inonog, S., Manoukian, A. S. & Young, M. W. A role for the segment polarity gene *shaggy/GSK-3* in the *Drosophila* circadian clock. *Cell* **105**, 769–779 (2001).
39. Vielhaber, E. L., Duricka, D., Ullman, K. S. & Virshup, D. M. Nuclear export of mammalian PERIOD proteins. *J. Biol. Chem.* **276**, 45921–45927 (2001).
40. Yagita, K. *et al.* Nucleocytoplasmic shuttling and mCRY-dependent inhibition of ubiquitylation of the mPER2 protein. *EMBO J.* **21**, 1301–1314 (2002).
41. McNamara, P., Seo, S.-B., Rudic, R. D., Sehgal, A., Chakravarti, D. & FitzGerald, G. A. Regulation of CLOCK and MOP4 by nuclear hormone receptors in the vasculature: a humoral mechanism to reset a peripheral clock. *Cell* **105**, 877–889 (2001).
42. Rutter, J., Reick, M., Wu, L. C. W. & McKnight, S. L. Regulation of CLOCK and NPAS2 DNA binding by the redox state of NAD cofactors. *Science* **293**, 510–514 (2001).
43. Zhang, Q., Piston, D. W. & Goodman, R. H. Regulation of corepressor function by nuclear NADH. *Science* **295**, 1895–1897 (2002).
44. Wright, K. P. Jr & Czeisler, C. A. Absence of circadian phase resetting in response to bright light behind the knee. *Science* **297**, 571 (2002).
45. Lucas, R. J. *et al.* Identifying the photoreceptive inputs to the mammalian circadian system using transgenic and retinally degenerate mice. *Behav. Brain Res.* **125**, 97–102 (2001).
46. Provencio, I. *et al.* A novel human opsin in the inner retina. *J. Neurosci.* **20**, 600–605 (2000).
47. Hattar, S., Liao, H. W., Takao, M., Berson, D. M. & Yau, K. W. Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* **295**, 1065–1070 (2002).
48. Provencio, I., Rollag, M. D. & Castrucci, A. M. Photoreceptive net in the mammalian retina. *Nature* **415**, 493 (2002).
49. Gooley, J. J., Lu, J., Chou, T. C., Scammell, T. E. & Saper, C. B. Melanopsin in cells of origin of the retinohypothalamic tract. *Nature Neurosci.* **4**, 1165 (2001).
50. Hannibal, J., Hindersson, P., Knudsen, S. M., Georg, B. & Fahrenkrug, J. The photopigment melanopsin is exclusively present in pituitary adenylate cyclase-activating polypeptide-containing retinal ganglion cells of the retinohypothalamic tract. *J. Neurosci.* **22**, RC191 (2002).
51. Berson, D. M., Dunn, F. A. & Takao, M. Phototransduction by retinal ganglion cells that set the circadian clock. *Science* **295**, 1070–1073 (2002).
52. Selby, C. P., Thompson, C., Schmitz, T. M., Van Gelder, R. N. & Sancar, A. Functional redundancy of cryptochromes and classical photoreceptors for nonvisual ocular photoreception in mice. *Proc. Natl Acad. Sci. USA* **97**, 14697–14702 (2000).
53. Thompson, C. L. *et al.* Preservation of light signaling to the suprachiasmatic nucleus in vitamin A-deficient mice. *Proc. Natl Acad. Sci. USA* **98**, 11708–11713 (2001).
54. Griffin, E. A., Staknis, D. & Weitz, C. J. Light-independent role of CRY1 and CRY2 in the mammalian circadian clock. *Science* **286**, 768–771 (1999).
55. Shigeyoshi, Y. *et al.* Light-induced resetting of a mammalian circadian clock is associated with rapid induction of the mPer1 transcript. *Cell* **91**, 1043–1053 (1997).
56. Albrecht, U., Zheng, B., Larkin, D., Sun, Z. S. & Lee, C. C. mPer1 and mPer2 are essential for normal resetting of the circadian clock. *J. Biol. Rhythms* **16**, 100–104 (2001).
57. Field, M. D. *et al.* Analysis of clock proteins in mouse SCN demonstrates phylogenetic divergence of the circadian clockwork and resetting mechanisms. *Neuron* **251**, 437–447 (2000).
58. Crosio, C., Cermakian, N., Allis, C. D. & Sassone-Corsi, P. Light induced chromatin modification in cells of the mammalian circadian clock. *Nature Neurosci.* **3**, 1241–1247 (2000).
59. Travnickova-Bendova, Z., Cermakian, N., Reppert, S. M. & Sassone-Corsi, P. Bimodal regulation of mPeriod promoters by CREB-dependent signaling and CLOCK:BMAL1 activity. *Proc. Natl Acad. Sci. USA* **99**, 7728–7733 (2002).
60. Shearman, L. P. & Weaver, D. R. Photic induction of *Period* gene expression is reduced in *Clock* mutant mice. *NeuroReport* **10**, 613–618 (1999).
61. Gau, D. *et al.* Phosphorylation of CREB Ser-142 regulates light-induced phase shifts of the circadian clock. *Neuron* **34**, 245–253 (2002).
62. Pennartz, J. A. M., de Jeu, M. T. G., Bos, N. P. A., Schaap, J. & Geurtsen, A. M. S. Diurnal modulation of pacemaker potentials and calcium current in the mammalian circadian clock. *Nature* **416**, 286–290 (2002).
63. Kramer, A. *et al.* Regulation of daily locomotor activity and sleep by hypothalamic EGF receptor signaling. *Science* **294**, 2511–2515 (2001).
64. Cheng, M. Y. *et al.* Prokineticin 2 transmits the behavioural circadian rhythm of the suprachiasmatic nucleus. *Nature* **417**, 405–410 (2002).
65. Bartness, T. J., Song, C. K. & Demas, G. E. SCN efferents to peripheral tissues: implications for biological rhythms. *J. Biol. Rhythms* **16**, 196–204 (2001).
66. Balsalobre, A. *et al.* Resetting of circadian time in peripheral tissues by glucocorticoid signaling. *Science* **289**, 2344–2347 (2000).
67. Damiola, F. *et al.* Restricted feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. *Genes Dev.* **14**, 2950–2961 (2000).
68. Stokkan, K. A., Yamazaki, S., Tei, H., Sakaki, Y. & Menaker, M. Entrainment of the circadian clock in the liver by feeding. *Science* **291**, 490–493 (2001).
69. Le Minh, N., Damilao, F., Tronche, F., Schutz, G. & Schibler, U. Glucocorticoids inhibit food-induced phase-shifting in peripheral circadian oscillators. *EMBO J.* **20**, 7128–7136 (2001).
70. Jin, X. *et al.* A molecular mechanism regulating rhythmic output from the suprachiasmatic circadian clock. *Cell* **96**, 57–68 (1999).
71. Ripperger, J. A. *et al.* CLOCK, an essential pacemaker component, controls expression of the circadian transcription factor DBP. *Genes Dev.* **14**, 679–689 (2000).
72. Lavery, D. J. *et al.* Circadian expression of the steroid 15 α -hydroxylase (*Cyp2a4*) and coumarin 7-hydroxylase (*Cyp2a5*) genes in mouse liver is regulated by the PAR leucine zipper transcription factor DBP. *Mol. Cell. Biol.* **19**, 6488–6499 (1999).
73. Mitsui, S., Yamaguchi, S., Matsuo, T., Ishida, Y. & Okamura, H. Antagonistic role of E4BP4 and PAR proteins in the circadian oscillatory mechanism. *Genes Dev.* **15**, 995–1006 (2001).
74. Clayton, J. D., Kyriacou, C. P. & Reppert, S. M. Keeping time with the human genome. *Nature* **409**, 829–831 (2000).
75. Grundschober, C. *et al.* Circadian regulation of diverse gene products revealed by mRNA expression profiling of synchronized fibroblasts. *J. Biol. Chem.* **276**, 46751–46758 (2001).
76. Duffield, G. E. *et al.* Circadian programs of transcriptional activation, signaling, and protein turnover revealed by microarray analysis of mammalian cells. *Curr. Biol.* **12**, 551–557 (2002).
77. Kita, Y. *et al.* Implications of circadian gene expression in kidney, liver and the effects of fasting on pharmacogenomic studies. *Pharmacogenetics* **12**, 55–65 (2002).
78. Akhtar, R. A. *et al.* Circadian cycling of the mouse liver transcriptome, as revealed by cDNA microarray, is driven by the suprachiasmatic nucleus. *Curr. Biol.* **12**, 540–550 (2001).
79. Panda, S. *et al.* Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* **109**, 307–320 (2002).
80. Storch, K. F. *et al.* Extensive and divergent circadian gene expression in liver and heart. *Nature* **417**, 78–83 (2002).
81. Harmar, A. J. *et al.* The VPAC₂ receptor is essential for circadian function in the mouse suprachiasmatic nuclei. *Cell* **109**, 497–508 (2002).
82. Reick, M., Garcia, J. A., Dudley, C. & McKnight, S. L. NPAS2: an analog of clock operative in the mammalian forebrain. *Science* **293**, 506–509 (2001).
83. Liu, C. & Reppert, S. M. GABA synchronizes clock cells within the suprachiasmatic circadian clock. *Neuron* **25**, 123–128 (2000).

Acknowledgements

We thank members of the Reppert laboratory for contributing to the ideas presented in this review, and U. Schibler for sharing data before publication. This work was supported by grants from the NIH and DARPA.

Correspondence and requests for materials should be addressed to S.M.R. (e-mail: steven.reppert@umassmed.edu).

Open channel structure of MscL and the gating mechanism of mechanosensitive channels

Eduardo Perozo*, D. Marien Cortes*, Pornthep Sompornpisut*†, Anna Kloda‡ & Boris Martinac‡

* Department of Molecular Physiology and Biological Physics, and Center for Structural Biology, University of Virginia, Charlottesville, Virginia 22906, USA

‡ Department of Pharmacology, Queen Elizabeth II Medical Center, University of Western Australia, Crawley, Western Australia 6009, Australia

Mechanosensitive channels act as membrane-embedded mechano-electrical switches, opening a large water-filled pore in response to lipid bilayer deformations. This process is critical to the response of living organisms to direct physical stimulation, such as in touch, hearing and osmoregulation. Here, we have determined the structural rearrangements that underlie these events in the large prokaryotic mechanosensitive channel (MscL) using electron paramagnetic resonance spectroscopy and site-directed spin labelling. MscL was trapped in both the open and in an intermediate closed state by modulating bilayer morphology. Transition to the intermediate state is characterized by small movements in the first transmembrane helix (TM1). Subsequent transitions to the open state are accompanied by massive rearrangements in both TM1 and TM2, as shown by large increases in probe dynamics, solvent accessibility and the elimination of all intersubunit spin-spin interactions. The open state is highly dynamic, supporting a water-filled pore of at least 25 Å, lined mostly by TM1. These structures suggest a plausible molecular mechanism of gating in mechanosensitive channels.

During severe osmotic challenges, the large-conductance mechanosensitive channel (MscL) of prokaryotes responds to changes in membrane tension by opening a wide water-filled pore that acts as an emergency safety valve^{1–3}. This physiological response is driven solely by changes in intrabilayer pressure gradients, a transduction mechanism which is common for all known prokaryotic and some (but not all) eukaryotic mechanosensitive channels^{4–7}.

A critical step towards understanding the molecular mechanisms of mechanosensation was recently realized when the structure of a member of the MscL family was solved at 3.5 Å resolution⁸. The structure shows that MscL is a highly α -helical homopentamer, where each subunit is composed of two transmembrane segments and a carboxy-terminal helical bundle protruding ~35 Å towards the cytoplasm. A number of gain-of-function mutants discovered at or near the amino-terminal end of TM1 point to the critical role that this region has during gating^{9–11}. In addition, *in vitro* sieving experiments¹² indicate that in the open state, the size of the open pore may reach 30–40 Å in diameter, perhaps one of the largest conformational changes known in membrane proteins.

In view of the specific physical constraints set by the crystal structure, a number of proposals have emerged for the molecular mechanism of gating in MscL. Some of these models^{8,10,12–14} are based on the observation that, given the limited number of TM segments and the experimental evidence for a large open pore, each helix must be positioned like the staves of a barrel in order to accommodate such large internal volume. In this way, residues from both TM segments would line the open pore. Recently, an alternative mechanism has been suggested on the basis of molecular modelling and crosslinking experiments^{15,16}. In this model, the channel opens by the widening of two gates in tandem: helical tilting and expansion of TM1, together with the separation of S1, a small helix located near the cytoplasmic membrane interface, and which is not resolved in the crystal structure.

We have pursued a spectroscopic approach to the structural analysis of gating of MscL in its native environment. The challenge

has been not only to obtain a stable open MscL under conditions of zero transbilayer pressures, but to do so in the entire vesicle population under observation. We found the solution to this problem to be the distortion of transbilayer pressure profiles through the use of mixtures of lipids with different geometries¹⁷. Previous experiments have shown that addition of charged amphiphiles or of lysophospholipids to bilayers containing mechanosensitive channels dramatically lowers their activation threshold^{4,18}. Using this approach, we demonstrated that MscL could indeed be trapped in the open state by the asymmetric incorporation of lysophosphatidylcholine (LPC) into phosphatidylcholine (PC) vesicles¹⁷. In addition, the channel could also be held in a structurally distinct intermediate conformation by reconstitution in thin dimyristoyl-phosphatidylcholine membranes (made of PC14 or PC16)¹⁷. The functional behaviour of MscL in its intermediate (PC16–PC14) and open conformations (LPC–PC mixture) is shown in Fig. 1a. Single-channel recordings reveal that only the curvature stress induced by LPC can trigger the transition to the fully open state in the absence of applied transbilayer pressures, while the intermediate state stabilized by thin bilayers is still non-conductive (although it requires lower pressures to open).

Here, we investigated the structural dynamic properties of these two conformations through site-directed spin-labelling (SDSL) and electron paramagnetic resonance (EPR) spectroscopy in relation to the known MscL structure in the closed state. This technique is powerful in the analysis of membrane protein structures and their conformational changes^{19–21}. A total of 55 cysteine mutants, targeted to the two TM segments (Fig. 1b), were generated and used to attach a nitroxide spin-label with a methanethiosulphonate moiety (Fig. 1c). Analysis of individual spectra provided information on probe dynamics, solvent accessibility and approximate intersubunit proximities in the conformationally trapped states. This information was compared with an earlier SDSL/EPR study of MscL²², to produce three-dimensional models of MscL at different stages of the gating process.

Transmembrane helical movements

The global differences in spin-label dynamics calculated from the

† Permanent address: Department of Chemistry, Faculty of Science, Chulalongkorn University, Phayathai Rd, Pratumwan, Bangkok 10330, Thailand.

entire EPR spectral data set are shown in Fig. 1d. Box plots for each of the TM segments represent the range of probe mobilities (ΔH_0^{-1}) in the closed state (reconstituted in PC18 membranes), the intermediate closed state (reconstituted in PC14 membranes) and in the fully open state (25 mol.% LPC in PC18). MscL shows similar overall dynamics in PC18 and PC14 for both TM1 (left panel) and TM2 (right panel), but undergoes a sharp increase in average mobility in the presence of LPC. This increase is most probably due to the significant reduction in interhelical packing that presumably accompanies the gating transition to the fully open state. A

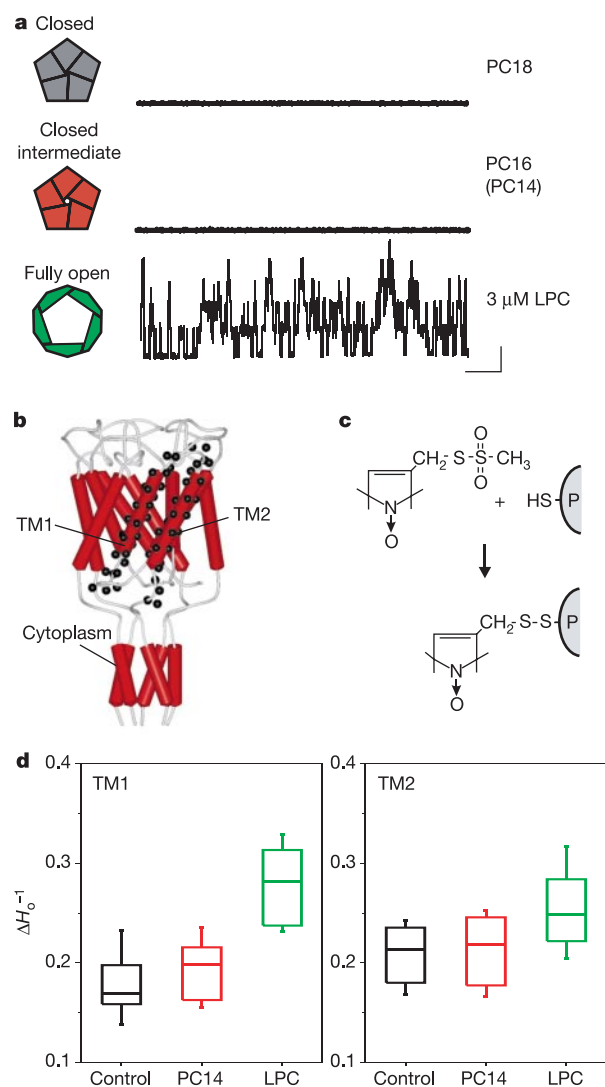


Figure 1 Functional and structural correlates of MscL gating. **a**, Experimental conditions used to trap different conformations of MscL: reconstitution into 18:1 dioleoyl-phosphatidylcholine (PC18) stabilizes the closed state (top); reconstitution into 14:1 dimiristoyl-phosphatidylcholine (PC14) stabilizes an intermediate closed conformation (centre); addition of lysophosphatidylcholine (LPC) locks MscL in the fully open conformation (bottom)¹⁷. Representative single-channel traces obtained in the absence of transbilayer pressure illustrate the functional state of the channel under conditions similar to those used for the EPR experiments. Horizontal scale: 5 s; vertical scale: 100 pA. **b**, Structure of MscL in the closed state. The amino-acid residues subjected to cysteine scanning mutagenesis in the present study are shown as black spheres. **c**, Chemical structure of the methanethiosulphonate spin label and its reaction with a protein sulphhydryl group. **d**, Box plots of the mobility parameter ΔH_0^{-1} taken over the whole set of transmembrane segments TM1 (left) or TM2 (right) EPR spectra. Each box represents 25 percentile, mean and 75 percentile of the data; the bars show the standard deviation from the mean.

detailed analysis of the structural changes in each conformation was carried out using three experimentally determined parameters that define the local environment surrounding a given spin label: (1) probe mobility (ΔH_0^{-1}); (2) collision with O_2 , which reflects accessibility to the membrane lipids ($\Delta P_{1/2}O_2$); and (3) collision with the nickel chelated complex Ni(II)-ethylenediaminediacetate (NiEdda), an indication of accessibility to the aqueous media ($\Delta P_{1/2}NiEdda$)^{19–21}. These parameters, when plotted as a function of residue number, represent a dynamic structural blueprint of MscL in a particular conformation.

In PC14 bilayers, a series of subtle structural rearrangements take place during the transition to the stable closed intermediate. In Fig. 2a, we focus on a nine-residue stretch at the core of the narrow pore constriction, the most obvious region to monitor gating transitions in MscL^{9,17}. Overall, the positions in this particular segment show only small changes in spectral line shape compared to those of the closed state, confirming that this region remains firmly immobilized²². However, rather significant spectral changes can be seen at residues I24 and A28, where an increase in spectral broadening suggests that some change in intersubunit distances takes place among equivalent positions. Residue environmental parameter profiles for TM1 and TM2 in the intermediate state (Fig. 2b, c) show that by far the largest changes occur in TM1. Changes in mobility and O_2 accessibility, and to a lesser extent a change towards the C-terminal end of the NiEdda accessibility profile, all point to a global rearrangement of the inner helical bundle. This includes an apparent ~ 40 -degree phase shift in TM1 that is most evident in the mobility profile (supporting a limited rotational helical movement, arrow). However, there is little or no change in $\Delta P_{1/2}NiEdda$ for the critical N-terminal half of TM1, which remains completely inaccessible to the aqueous milieu.

We found almost no spectral changes along TM2. In fact, all three residue environmental parameters appear to be in precise register with those in the closed conformation, an indication that TM2 does not rotate about its main axis in this intermediate state. Mapping the relative changes in the individual environmental parameters onto a cross-section of the closed MscL provides a structural context to the conformational changes in PC14, and confirms that the bulk of the structural rearrangements is limited to the channel core (Fig. 2d).

In dramatic contrast, stabilization of the open state using asymmetric mixtures LPC and PC18 revealed major structural changes throughout the MscL structure (Fig. 3). We found an almost uniform sharpening of spectral line shapes on the nine-residue section forming the hydrophobic pore constriction in the closed state, reflecting significant increases in probe dynamics (Fig. 3a). In TM1 (Fig. 3b), transition to the open state results in a widespread increase in probe mobility (ΔH_0^{-1} , top panel), a rather uniform increase in O_2 accessibility ($\Delta P_{1/2}O_2$, central panel) and a periodic NiEdda accessibility profile throughout the length of TM1 ($\Delta P_{1/2}NiEdda$, lower panel). The most important result, however, is the finding that even residues at the narrow constriction of the pore (horizontal bar in the lower panel of Fig. 3b) now seem to be fully exposed to NiEdda, a direct demonstration of TM1 participation in the conductive pathway.

Although local environment changes in TM2 (Fig. 3c) appear to be a bit more limited, we note however that these are not expected to be large for an exposed helix that is surrounded by lipids in both closed and open conformations. In fact, the points at which we see the smallest change in ΔH_0^{-1} (arrows in top panel of Fig. 3c) precisely correspond to putatively lipid-exposed residues in the closed structure⁸. We found that with the exception of the first helical turn at the extracellular end of TM2 (residues 72–75), there is essentially no change in the accessibility to NiEdda (Fig. 3c, lower panel). This result is inconsistent with gating models in which the lining of the open pore is somehow shared between the two TM segments acting as ‘barrel staves’^{8,10,12–14}. Mapping the environmen-

tal parameters onto the closed structure surface corroborates that a much larger and widespread structural rearrangement takes place in the open state (Fig. 3d). This change requires not only significant interhelical separations (arrows), but as in the case of $\Delta P_{1/2}\text{NiEdda}$ in TM1, substantial rotations about its principal axis.

Analysis of the changes in spin–spin interaction in both helices offers a more specific method of monitoring of intersubunit separation (Fig. 4). As a five-fold symmetric homopentamer, fully labelled MscL is a multi-spin system in which singly labelled subunits show strong spin–spin interaction for residues near the symmetry axis. Thus, any protein movement that alters the position of a residue relative to the symmetry axis can generate changes in spin–spin coupling (qualitatively estimated as the Ω parameter) as long as the spin labels are within ~ 15 Å. Figure 4a shows this to be the case for a number of TM1 residues in the PC14 intermediate conformation. Although the pattern is complex, it is clear that in PC14 membranes the centre of the helix gets closer to the symmetry axis, while slightly separating the two ends of the helix. More importantly, unlike the observations on the closed and intermediate states, we found no evidence of spectral broadening due to intersubunit spin–spin interactions in the open state. Thus, in the open state the TM segments move so that the positions of the individual spins lie beyond ~ 15 Å from each other, as expected from the generation of a large pore upon gating (Fig. 4b).

Restraint-driven structure modelling

The present data set offers a fairly extensive collection of constraints that can be used to derive structural models of MscL in the open and intermediate states. However, because of the magnitude of the conformational rearrangements in the open state, the available data is mostly limited to changes in local environmental parameters, with few reliable intersubunit distance constraints. We have met this challenge through the use of a simple and efficient computational method based on the exhaustive sampling of rigid-body movements in cartesian space (restraint-driven cartesian transformation, ReD-CaT)²³. This approach has been used to determine the type, direction and magnitude of the conformational changes in a defined secondary structure element using limited structural information²⁴.

Model building was restricted to the intramembrane regions of MscL and therefore included no information on the extracellular loops, the cytoplasmic five-helix bundle at the C terminus, or the N-terminal end. Starting with the closed-state model, each TM segment was allowed five degrees of freedom and subjected to geometric transformations in cartesian space. Conformational space was systematically searched by manipulating rotations and lateral displacements of individual TM segments, relative to x -, y -, z -, and the principal helical axis. In each evaluation step, a penalty function was calculated using changes in the solvent accessibility between a given state and the closed structure. Restraints used by

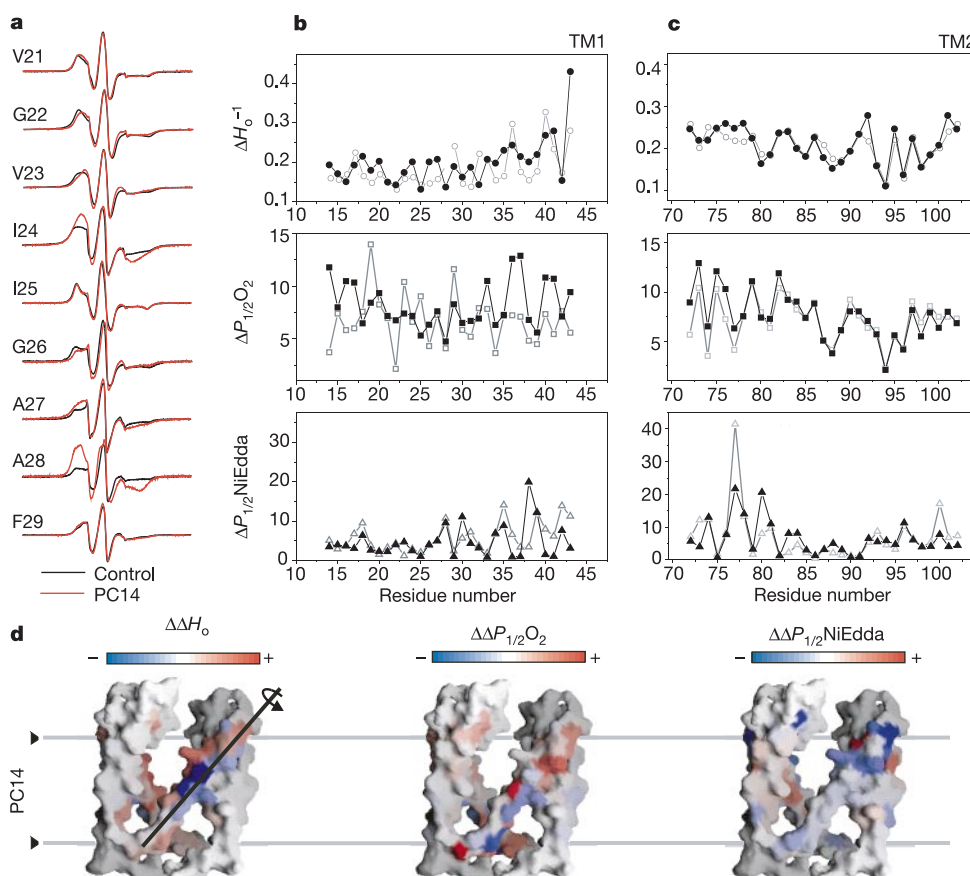


Figure 2 Structural rearrangements underlying the transition to the PC14 intermediate state. **a**, CW-EPR spectra of residues forming the narrowest portion of the permeation path in TM1 (residues 21–29). Overlapping spectra are shown for channels in the closed conformation (black traces) and the intermediate state (red traces). Spectral width is 150 G in all cases. **b**, **c**, Residue-specific environmental parameter profiles obtained in PC14 vesicles for transmembrane domains TM1 (residues 14–43, **b**) and TM2 (residues 72–96, **c**): mobility parameter ΔH_0^{-1} (top panels, filled circles), oxygen accessibility

parameter $\Delta P_{1/2}\text{O}_2$ (middle panels, filled squares), and NiEdda accessibility parameter $\Delta P_{1/2}\text{NiEdda}$ (bottom panels, filled triangles). In each panel, the corresponding open symbol plot represents data obtained in the closed state²². **d**, Extent of environmental parameter changes ($\Delta\Delta H_0^{-1}$, $\Delta\Delta P_{1/2}\text{O}_2$ and $\Delta\Delta P_{1/2}\text{NiEdda}$) mapped onto a molecular surface rendering of the closed conformation (two subunits are shown for clarity). Molecular surface and data mapping was done using the program GRASP³⁸.

ReDCaT were derived from the normalized $\Delta\Delta P_{1/2}$ of O₂ and NiEdda, and the calculated accessibility delta was obtained from the solvent accessible surface areas²⁵ of sampling models subtracted from those of the closed MscL structure, and subsequently normalized (see Methods).

Structure of the PC14 intermediate

Figure 5 shows the structural model of MscL in its PC14 closed intermediate state. As expected, the conformational changes are small and restricted to the core helix, TM1. Indeed, given the minute changes in the spectral properties observed for TM2, the coordinates of this helix must remain unchanged from those in the closed state. Figure 5a shows a comparison of the structural rearrangements in TM1 between the closed (green model) and the PC14 intermediate states (red model). The helices are shown within a solvent-accessible surface calculated with a 2 Å probe. The PC14 intermediate state is characterized by a clockwise rotation about the main helical axis of about $\sim 30^\circ$, with a small anticlockwise tilt in the *z* axis and a slight tilt towards the *x-y* plane (the bilayer). This helical movement narrows the pore at the base of the external vestibule by about 1–2 Å while slightly increasing the interhelical separation towards the N-terminal end of TM1.

Structure of the open state

The open-state model (Fig. 6) was generated almost solely on the basis of accessibility data. However, the magnitude of the structural

changes is such that independent ReDCaT runs tended to converge on the same type of structure. Figure 6a offers two views of open MscL, each subunit colour-coded and shown in the context of a 4 Å probe solvent-accessible surface. The most salient feature of the open model is the presence of a large (≥ 25 Å), water-filled pore lined almost exclusively by TM1. This helix in turn is shielded from extensive lipid contacts by the TM2 segment from a neighbouring subunit. These pore dimensions would be large enough to support a nanoSiemens level conductance, linearly dependent with the conductivity of the solution^{3,26,27}. To generate such a large pore, both TM segments appear significantly tilted in relation with the membrane normal ($\sim 45^\circ$, on average), and as a consequence, the channel flattens considerably in the projection to the membrane normal (≥ 25 Å) but expands in the membrane plane (to ≥ 70 Å). This phenomenon has also been observed in simulations of MscL gating^{16,28}. In principle, this would necessitate important physical rearrangements in the surrounding lipid, but at present this prediction is difficult to assess in the absence of structural data for the channel regions located outside the membrane. A visual correlation between these structural models and the EPR data is shown in the Supplementary Information.

Molecular mechanism of gating in MscL

With backbone-level models for the transmembrane helices in the closed, intermediate and fully open states, we can now examine the sequence of molecular rearrangements taking place during MscL

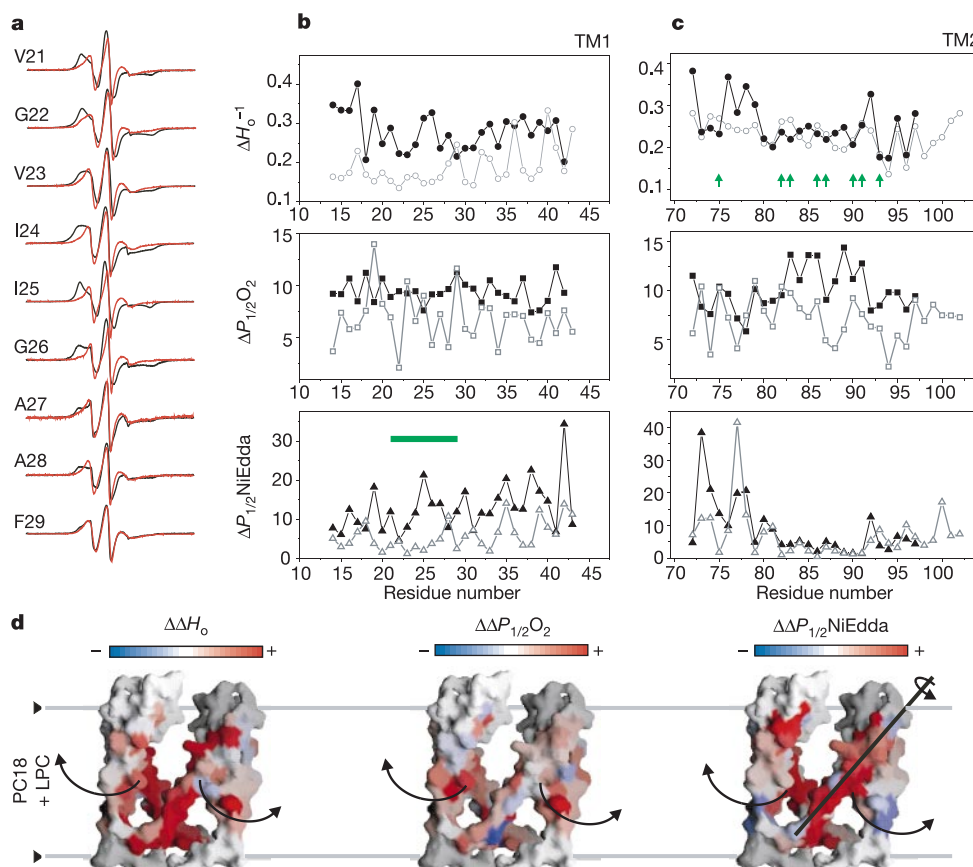


Figure 3 Structural rearrangements underlying the transition to the open state.

a, Continuous wave EPR spectra of residues forming the narrowest portion of the permeation path in TM1 (residues 21–29). Black traces were obtained from channels in the closed conformation and red traces represent the open state. Spectral conditions identical to Fig. 2. **b**, **c**, Residue-specific environmental parameter profiles obtained in PC18 + 25 mol.% LPC vesicles for transmembrane domains TM1 (**b**) and TM2 (**c**): ΔH_0^{-1} (top panels, filled circles), $\Delta P_{1/2}O_2$ (middle panels, filled squares), and

$\Delta P_{1/2}NiEdda$ (bottom panels, filled triangles). The green bar highlights NiEdda accessibility changes at the TM1 pore constriction. Green arrows identify residues in TM2 that are fully exposed to lipids in the closed state. Again, the corresponding open symbol plot represents data obtained in the closed state²². **d**, Extent of environmental parameter changes mapped onto a molecular surface rendering of the closed conformation. Details identical to those in Fig. 2.

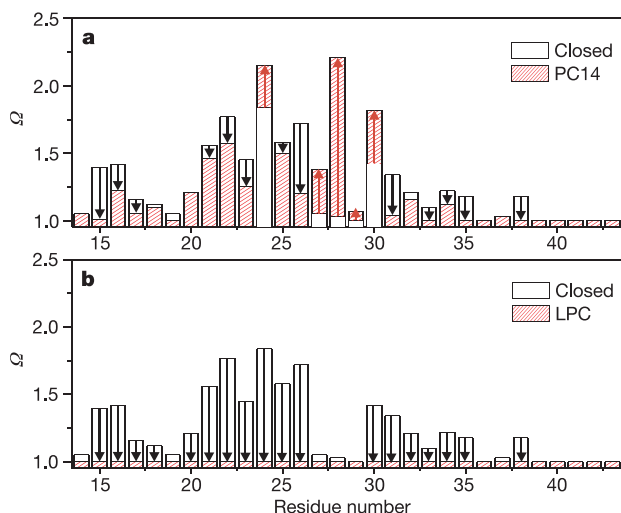


Figure 4 Changes in the pattern of TM1 intersubunit proximities upon transition to the intermediate and open states. Open bars represent Q values obtained in the closed state (PC18), while patterned bars correspond to data in the intermediate state (**a**) and the open state (**b**). Arrows show the direction of change: downward arrows (black) point to a reduction in spin–spin coupling (residues are moving away from each other and the symmetry axis).

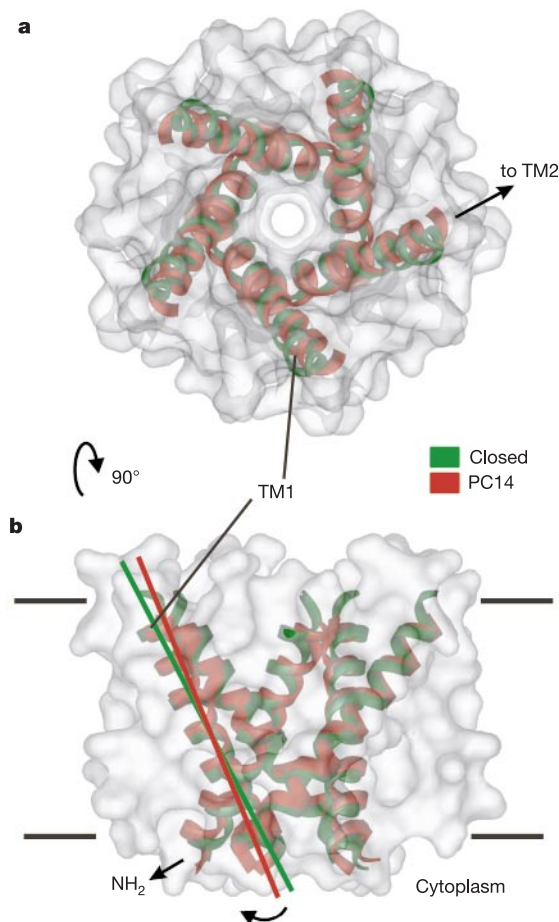


Figure 5 Structure of MscL transmembrane segments in the intermediate closed state (PC14). Top (**a**) and side (**b**) views of TM1 in the intermediate closed pentamer (red ribbon), compared to the crystal structure (green ribbon). Structures are shown within a translucent molecular surface representation of all TM segments.

gating (Fig. 7). The early structural changes associated with PC14 membranes are limited to a small set of rotations and tilts in TM1 (Fig. 7a); however, these are likely to represent a substantial fraction of the energetic cost of opening MscL, as shown by the significant shifts in pressure activation seen during patch clamp experiments in thin bilayers¹⁷. Further along the path towards the open state, TM1 continues to rotate along its principal axis (some additional 60–70°), and tilts about 15° towards the bilayer plane while shifting away from the symmetry axis at least 13 Å. TM2, on the other hand, moves perpendicularly to TM1 while tilting some 17° towards the bilayer and rotating at least 70° in the anticlockwise direction. These values are summarized in the Supplementary Information. We stress, nonetheless, that these values should be interpreted with care, as the model reflects only a direct comparison between TM segments, without considering the contributions of the extra-membrane elements. Vertical cross-sections of the permeation pathway calculated using the HOLE program²⁹ for the three stable conformers highlight the morphological changes in the permeation pathway. In PC14 membranes, there is an initial widening of the pore of about ~5 Å towards the cytoplasmic side of the channel, hinting at an intracellular initiation of the massive opening that leads to the open pore.

The general features of the present open channel structure are in agreement with earlier computer-generated models of the open state^{15,16}. In particular, there are similarities regarding the sharp increase in helical tilting towards the plane of the bilayer and the spatial relation between TM1, TM2 and the aqueous pore. However, the actual path that the channel follows from the closed state will

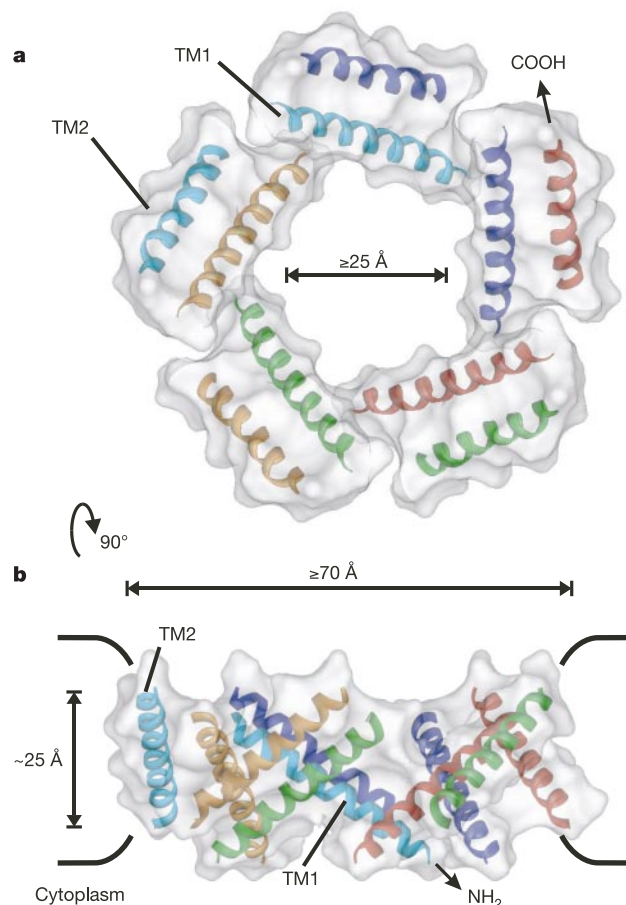


Figure 6 Structure of MscL TM segments in the open state. Extracellular (**a**) and side (**b**) views of TM1 and TM2 in the open state. The side view is shown in relation to a hypothetically distorted bilayer. TM segments of the same subunit have the same colour.

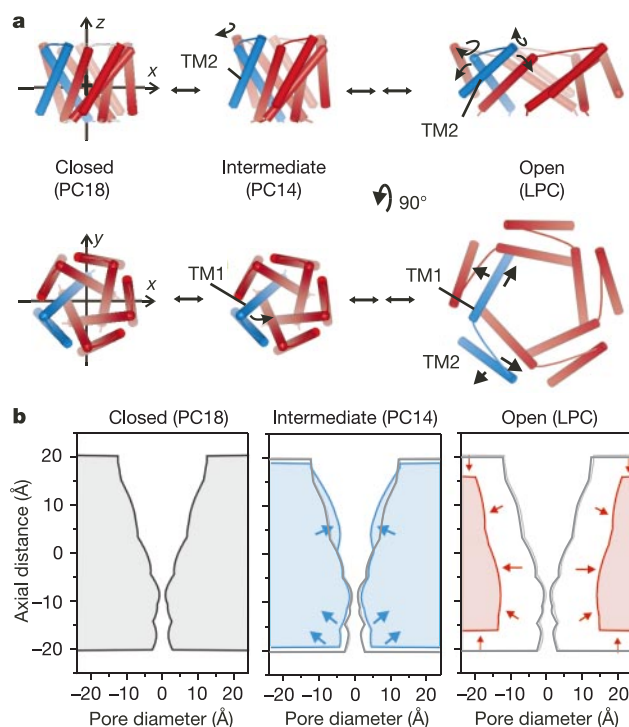


Figure 7 A structural model of gating in MscL. **a**, Side (top) and extracellular (bottom) view of the sequence of structural rearrangements leading to the open state. The axes of cartesian space are shown on the closed MscL structure. Helical movements are illustrated by black arrows on a single MscL subunit highlighted in blue. **b**, Cross-sections of the MscL pore in the closed (left), intermediate (centre) and open (right) conformations. As a reference, the closed state is present in all panels (thin line). Each cross-section was obtained from the calculated pore surface using the program HOLE²⁹. The intermediate and open cross-sections are shown in relation to the one in the closed state (outlined pore surface).

require further examination. If MscL in PC14 membranes represents an early intermediate along the reaction coordinate towards the open state¹⁷, the gating mechanism suggested by the present set of structures does not seem to support the existence of an early expansion step preceding the fully open state. However, it is still possible that PC14 may be stabilizing an intermediate through a kinetic pathway independent of that leading to the LPC-stabilized open state. Nonetheless, the large structural changes that occur between the PC14 intermediate and the open state call for several additional intermediate states. Included among them are possible silent expansions as well as the multiple subconducting states detected using patch clamp methods³⁰. □

Methods

Generation of spin-labelled mutants

Cysteine mutants were generated by oligonucleotide mismatch site-directed mutagenesis using the Transformer kit (Clontec) as previously described^{17,22}. Individual mutants were expressed, purified, spin-labelled and reconstituted following published protocols^{17,22}. Briefly, N-terminal histidine-tagged mutants were purified on a Co²⁺-based metal-chelate chromatography resin, labelled twice at a 10:1 label:channel molar ratio with methanethiosulphonate spin label (Toronto Research). Underlabelled channels were incubated once at a 1:10 ratio. All samples were divided in two and reconstituted in either PC14 or PC18 membranes.

Electron paramagnetic resonance spectroscopy and analysis

Details of the EPR spectroscopy methods have been published elsewhere^{31,32}. All X-band continuous wave (CW) EPR spectra were recorded in a Bruker EMX spectrometer fitted with a loop-gap resonator under the following conditions: 2 mW incident power, 100 kHz modulation frequency and 1 G modulation amplitude. Solvent accessibility was estimated from the difference in spectral saturation between a sample of a test solution (containing a paramagnetic collisional probe) and the same sample in the presence of molecular nitrogen (no effective collisions)³³. Power saturation curves were obtained for each spin-

labelled mutant after equilibration in N₂ as control, and air (21% O₂), and N₂ in the presence of 200 mM NiEdda as relaxing agents. All EPR data were obtained at room temperature.

When the interspin vector changes dynamically on the EPR timescale, the static dipolar interaction is partially averaged owing to homogeneous broadening. This results in a relaxation mechanism that imposes a practical upper limit of ~15 Å to the interspin distance at which broadening derived from spin-spin coupling can be detected at room temperature³⁴. Therefore, when compared to the spectral sum of the single mutants (no spin-spin coupling), the broadening of the EPR spectrum can be interpreted semi-quantitatively to define distances into three broad categories: close (<10 Å), intermediate (10–15 Å) and far (beyond 15 Å). Given the uncertainty due to the pentameric geometry of intersubunit distances, we used a qualitative estimate of intersubunit proximity, the Ω parameter. Ω is operationally defined from the normalized amplitude ratio between the under-labelled and fully labelled spectra³¹. Thus, at large interspin separations, the Ω parameter is approximately one (no spin-spin coupling), but Ω values increase as the spin-labels become closer to each other.

Structure modelling

ReDCaT, a constraint-biased sampling of conformational space in a cartesian coordinate system, was used to select the configurations that best fit the individual EPR data set in the intermediate and the open states with the closed MscL structure as a reference³⁵. An initial homology model of Eco-MscL was generated for TM1 and TM2 based on the Tb-MscL structure (PDB accession code 1MSL). We used a five-fold rotational symmetric operation on both segments to obtain structural coordinates of the homo-pentameric protein. The model was limited to residues 14–43 for TM1 and 72–96 for TM2, with all TM domains built as ideal α -helices. The original ReDCaT algorithm was modified to include, in the penalty calculations, contributions from solvent accessibility deltas of the form:

$$\Delta A_{\text{EPR}} = A_{\text{EPR}}(\text{Open, LPC}) - A_{\text{EPR}}(\text{Closed})$$

$$\text{or } \Delta A_{\text{EPR}} = A_{\text{EPR}}(\text{Intermediate, PC14}) - A_{\text{EPR}}(\text{Closed})$$

where A_{EPR} is the EPR-derived accessibility parameter calculated from the normalized changes in O₂ and NiEdda accessibilities—norm $\Delta\Delta P_{1/2\text{O}_2}$ and norm $\Delta\Delta P_{1/2\text{NiEdda}}$, respectively. Upper and lower bound restraints were defined as:

$$\Delta A_{\text{EPR}}^{\text{upper}} = \Delta A_{\text{EPR}} + 0.3\Delta A_{\text{EPR}}$$

$$\Delta A_{\text{EPR}}^{\text{lower}} = \Delta A_{\text{EPR}} - 0.3\Delta A_{\text{EPR}}$$

The bounds range was defined by taking 30% deviation of ΔA_{EPR} ²³ and from initial experimental correlations between experimental and calculated solvent accessibilities³⁵. Comparison of the experimental O₂ and NiEdda accessibility data to the calculated data was carried out according to ref. 25 using probe radii of 2 Å and 4 Å respectively. Furthermore, for the intermediate structure model, a ± 2 Å cutoff distance change criterion was defined among adjacent TM1 subunits based on the change in TM1 spectral broadening for the spin-labelled mutants reconstituted in PC14 relative to the control, screening the best-fit structures before penalty evaluation. For the open state, only a 15 Å cutoff distance between intersubunit residues of both TM1 and TM2 segments was used. An ensemble of structures for the intermediate and open states were selected on the basis of penalty value. The mean structures of the closed intermediate and the open state were subjected to energy minimization using Amber³⁶ with full atom representation. Protein structure analysis was carried out using Procheck³⁷, GRASP³⁸, HOLE²⁹ and WebLab viewer (Accelrys) were used for visualization and analysis. We note, however, that given the limited availability of well-defined distances and the reliance on solvent accessibility changes, the present structural models represent, at best, an approximation to the actual structure and should be interpreted with the proper care.

Received 4 April; accepted 15 July 2002; doi:10.1038/nature00992.

1. Sukharev, S. I., Blount, P., Martinac, B. & Kung, C. Mechanosensitive channels of *Escherichia coli*: the MscL gene, protein, and activities. *Annu. Rev. Physiol.* **59**, 633–657 (1997).
2. Wood, J. M. Osmosensing by bacteria: signals and membrane-based sensors. *Microbiol. Mol. Biol. Rev.* **63**, 230–262 (1999).
3. Hamill, O. P. & Martinac, B. Molecular basis of mechanotransduction in living cells. *Physiol. Rev.* **81**, 685–740 (2001).
4. Martinac, B., Adler, J. & Kung, C. Mechanosensitive ion channels of *E. coli* activated by amphipaths. *Nature* **348**, 261–263 (1990).
5. Sukharev, S. I., Blount, P., Martinac, B., Blattner, F. R. & Kung, C. A large-conductance mechanosensitive channel in *E. coli* encoded by *mscL* alone. *Nature* **368**, 265–268 (1994).
6. Patel, A. J. et al. A mammalian two pore domain mechano-gated S-like K⁺ channel. *EMBO J.* **17**, 4283–4290 (1998).
7. Levina, N. et al. Protection of *Escherichia coli* cells against extreme turgor by activation of MscS and MscL mechanosensitive channels: identification of genes required for MscS activity. *EMBO J.* **18**, 1730–1737 (1999).
8. Chang, G., Spencer, R. H., Lee, A. T., Barclay, M. T. & Rees, D. C. Structure of the MscL homolog from *Mycobacterium tuberculosis*: a gated mechanosensitive ion channel. *Science* **282**, 2220–2226 (1998).
9. Ou, X., Blount, P., Hoffman, R. J. & Kung, C. One face of a transmembrane helix is crucial in mechanosensitive channel gating. *Proc. Natl Acad. Sci. USA* **95**, 11471–11475 (1998).
10. Yoshimura, K., Batiza, A., Schroeder, M., Blount, P. & Kung, C. Hydrophilicity of a single residue within MscL correlates with increased channel mechanosensitivity. *Biophys. J.* **77**, 1960–1972 (1999).
11. Yoshimura, K., Batiza, A. & Kung, C. Chemically charging the pore constriction opens the mechanosensitive channel MscL. *Biophys. J.* **80**, 2198–2206 (2001).
12. Cruickshank, C. C., Minchin, R. F., Le Dain, A. C. & Martinac, B. Estimation of the pore size of the large-conductance mechanosensitive ion channel of *Escherichia coli*. *Biophys. J.* **73**, 1925–1931 (1997).
13. Spencer, R. H., Chang, G. & Rees, D. C. 'Feeling the pressure': structural insights into a gated

- mechanosensitive channel. *Curr. Opin. Struct. Biol.* **9**, 448–454 (1999); erratum **9**, 650–651 (1999).
14. Batiza, A. F., Rayment, I. & Kung, C. Channel gate: Tension, leak and disclosure. *Struct. Fold. Design* **7**, R99–R103 (1999).
15. Sukharev, S., Betanzos, M., Chiang, C.-S. & Guy, H. The gating mechanism of the large mechanosensitive channel MscL. *Nature* **409**, 720–724 (2001).
16. Sukharev, S., Durell, S. R. & Guy, H. R. Structural models of the MscL gating mechanism. *Biophys. J.* **81**, 917–936 (2001).
17. Perozo, E., Kloda, A., Cortes, D. M. & Martinac, B. Physical principles underlying the transduction of bilayer deformation forces during mechanosensitive channel gating. *Nature Struct. Biol.* **9**, 696–703 (2002).
18. Maingret, F., Patel, A. J., Lesage, F., Lazdunski, M. & Honoré, E. Lysophospholipids open the two-pore domain mechano-gated K⁺ channels TREK-1 and TRAAK. *J. Biol. Chem.* **275**, 10128–10133 (2000).
19. Hubbell, W. L., Gross, A., Langen, R. & Lietzow, M. A. Recent advances in site-directed spin labeling of proteins. *Curr. Opin. Struct. Biol.* **8**, 649–656 (1998).
20. Hubbell, W. L., Cafiso, D. S. & Altenbach, C. Identifying conformational changes with site-directed spin labeling. *Nature Struct. Biol.* **7**, 735–739 (2000).
21. Mchaourab, H. & Perozo, E. in *Biological Magnetic Resonance* Vol. 19 (eds Eaton, G. R., Eaton, S. S. & Berliner, L. J.) 155–218 (Kluwer-Plenum, New York, 2000).
22. Perozo, E., Kloda, A., Cortes, D. M. & Martinac, B. Site-directed spin-labeling analysis of reconstituted MscL in the closed state. *J. Gen. Physiol.* **118**, 193–206 (2001).
23. Sompornpisut, P., Liu, Y. S. & Perozo, E. Calculation of rigid-body conformational changes using restraint-driven Cartesian transformations. *Biophys. J.* **81**, 2530–2546 (2001).
24. Liu, Y. S., Sompornpisut, P. & Perozo, E. Structure of the KcsA channel intracellular gate in the open state. *Nature Struct. Biol.* **8**, 883–887 (2001).
25. Lee, B. & Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400 (1971).
26. Martinac, B., Buechner, M., Delcour, A. H., Adler, J. & Kung, C. Pressure-sensitive ion channel in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **84**, 2297–2301 (1987).
27. Martinac, B. Mechanosensitive channels in prokaryotes. *Cell Physiol. Biochem.* **11**, 61–76 (2001).
28. Gullingsrud, J., Kosztin, D. & Schulten, K. Structural determinants of MscL gating studied by molecular dynamics simulations. *Biophys. J.* **80**, 2074–2081 (2001).
29. Smart, O. S., Neduvellil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14**, 354–360 (1996).
30. Sukharev, S. I., Sigurdson, W. J., Kung, C. & Sachs, F. Energetic and spatial parameters for gating of the bacterial large conductance mechanosensitive channel, MscL. *J. Gen. Physiol.* **113**, 525–540 (1999).
31. Perozo, E., Cortes, D. M. & Cuello, L. G. Three-dimensional architecture and gating mechanism of a K⁺ channel studied by EPR spectroscopy. *Nature Struct. Biol.* **5**, 459–469 (1998).
32. Cortes, D. M., Cuello, L. G. & Perozo, E. Molecular architecture of full-length KcsA: role of cytoplasmic domains in ion permeation and activation gating. *J. Gen. Physiol.* **117**, 165–180 (2001).
33. Altenbach, C., Marti, T., Khorana, H. G. & Hubbell, W. L. Transmembrane protein structure: spin labeling of bacteriorhodopsin mutants. *Science* **248**, 1088–1092 (1990).
34. Abragam, A. *Principles of Nuclear Magnetism* (Oxford Univ. Press, Oxford, 1961).
35. Sompornpisut, P., Mchaourab, H. & Perozo, E. Spectroscopically-determined solvent accessibilities as constraints for global fold discrimination in proteins. *Biophys. J.* **82**, 474a (2002).
36. Case, D. A. *et al.* AMBER 6 (Univ. California, San Francisco, 1999).
37. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).
38. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements

We thank R. Nakamoto and M. Wiener for thoughtful discussions, C. Ptak for reading the manuscript, and S. Mochel for encouragement. This work was supported in part by the NIH (E.P.) and the McKnight endowment fund for neuroscience (E.P.), the Australian Research Council grants (B.M.) and the Australian Academy of Science (Scientific Visit Award to B.M.). Transmembrane segment coordinates in the three conformations have been deposited in the Protein Data Bank (accession codes 1KYK, 1KYL and 1KYM).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.P. (e-mail: eperozo@virginia.edu).

A short timescale for terrestrial planet formation from Hf–W chronometry of meteorites

Qingzhu Yin*, S. B. Jacobsen*, K. Yamashita*†, J. Blichert-Toft‡, P. Télouk‡ & F. Albarède‡

* Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, Massachusetts 02138, USA

‡ Laboratoire des Sciences de la Terre, Ecole Normale Supérieure de Lyon, 46, Allée d'Italie, 69364 Lyon Cedex 7, France

Determining the chronology for the assembly of planetary bodies in the early Solar System is essential for a complete understanding of star- and planet-formation processes. Various radio-nuclide chronometers (applied to meteorites) have been used to determine that basaltic lava flows on the surface of the asteroid Vesta formed within 3 million years (3 Myr) of the origin of the Solar System^{1–3}. Such rapid formation is broadly consistent with astronomical observations of young stellar objects, which suggest that formation of planetary systems occurs within a few million years after star formation^{4,5}. Some hafnium–tungsten isotope data, however, require that Vesta formed later⁶ (~16 Myr after the formation of the Solar System) and that the formation of the terrestrial planets took a much longer time^{7–10} (62^{+4504}_{-14} Myr). Here we report measurements of tungsten isotope compositions and hafnium–tungsten ratios of several meteorites. Our measurements indicate that, contrary to previous results^{7–10}, the bulk of metal–silicate separation in the Solar System was completed within <30 Myr. These results are completely consistent with other evidence for rapid planetary formation^{1–5}, and are also in agreement with dynamic accretion models^{11–13} that predict a relatively short time (~10 Myr) for the main growth stage of terrestrial planet formation.

It has been argued that the rate of terrestrial core formation was limited by accretion, that it started very early, and that it was largely completed within the first 10–20 Myr or less of Earth history^{14,15}. The decay of now extinct ¹⁸²Hf (half-life, 9 Myr) to ¹⁸²W is an ideal chronometer for tracing this process, because Hf is retained in the silicate mantle while W is largely partitioned into the core during core segregation^{6–10,14–20}. However, it is now widely accepted that the reports of almost identical ¹⁸²W/¹⁸³W ratios of chondritic meteorites and the bulk silicate Earth (BSE) imply that core segregation in the Earth took place about 60 Myr after the beginning of the Solar System^{7,8,10}—that is, when ¹⁸²Hf was essentially extinct. As core segregation probably occurred in an early terrestrial magma ocean, the observed solar He and Ne components of the Earth's mantle also require that most of the planetary accretion occurred before the gas of the solar nebula had dissipated^{21,22}. The timescale for this dissipation is <3 Myr after the beginning of the Solar System^{4,23}, and highlights the problem of inconsistent chronologies for the early Earth.

Our new data (Table 1 and ref. 24) for ordinary chondrites Dalgaty Downs and Dhurmsala yield a well defined 'fossil isochron' array (Fig. 1). Four separate whole-rock samples of carbonaceous chondrites (three from Allende and one from Murchison) and a calcium–aluminium–rich inclusion (CAI) from the Allende meteorite plot on this isochron. Although the absolute ages of the two ordinary chondrites are not well known independently, the fact that a CAI from the Allende meteorite plots within error of this isochron requires the Hf–W age of these ordinary chondrites to be within ~1

to 2 Myr of the age of Allende CAIs (the oldest solid objects of the Solar System). The chondrite isochron gives an initial ¹⁸²Hf/¹⁸⁰Hf ratio of $(1.00 \pm 0.08) \times 10^{-4}$ and an initial ¹⁸²W/¹⁸³W ratio corresponding to $\epsilon_w = -3.45 \pm 0.25$ (ϵ_w is defined in Table 1). The lowest initial ϵ_w we have obtained based on metals from chondrites and iron meteorites is about -3.5 ± 0.5 (Table 1 and unpublished data). Thus, the ¹⁸²Hf/¹⁸⁰Hf ratio of 1×10^{-4} and the initial ϵ_w of -3.5 must be very close to their respective initial solar values. This initial ¹⁸²Hf/¹⁸⁰Hf value is significantly lower than, and clearly inconsistent with, the commonly cited value of 2.75×10^{-4} (ref. 18). In addition to obtaining a lower initial solar ¹⁸²Hf abundance compared to the earlier work, one of the features of Fig. 1 is that three repeat measurements of Allende whole rock and the Dalgaty Downs whole rock all plot consistently ~2 ϵ units below the terrestrial standard and have a chondritic Hf/W ratio ($f^{\text{Hf/W}} = 0$; $f^{\text{Hf/W}}$ is defined in Table 1). This is inconsistent with previously measured whole-rock samples of Allende and Murchison^{7,8} that yielded $\epsilon_w \approx 0$ (discrepancies with published chondrite data are further discussed in Fig. 1 legend). Both ordinary chondrite isochrons are consistent with a chondritic uniform reservoir (CHUR) value of $\epsilon_w = -1.9 \pm 0.2$ and $f^{\text{Hf/W}} = 0$. This means that the BSE with its high Hf/W ratio has a radiogenic ¹⁸²W/¹⁸³W signature about 2 ϵ units higher than the chondritic value. A comparison of results based on the old and the new data are given in Table 2.

Our Juvinas (eucrite) measurement plots on the eucrite isochron of ref. 6 with ¹⁸²Hf/¹⁸⁰Hf = 7.96×10^{-5} . The eucrites post-date our ordinary chondrites by only ~3 Myr. In contrast, the Hf–W data for eucrites⁶, when compared to the currently accepted high initial solar ¹⁸²Hf/¹⁸⁰Hf value^{7,8,18,19}, indicate a late parent body (Vesta) differentiation at ~16 Myr after the formation of the Solar System. This is

Table 1 Hf–W isotope data for meteorites

Sample	Hf (p.p.b.)	Hf/W	$f^{\text{Hf/W}}$	ϵ_w	$\pm 2\sigma$
Allende, WR no. 1	187.5	1.096	-0.0236	-1.93	0.28
Allende, WR no. 2	187.7	1.098	-0.0226	-2.03	0.82
Allende, WR no. 3	210.2	1.229	0.0946	-2.02	0.36
Allende, CAI (A44A)	5,639	3.477	2.095	1.16	0.55
Murchison, WR	157.1	0.878	-0.218	-2.62	0.33
Dhurmsala, silicate	185.2	6.717	4.982	+5.81	0.67
Dhurmsala, metal	8.25	0.0382	-0.966	-3.07	0.35
Dhurmsala, WVR	181.1	1.395	0.242	-1.27	0.51
Dalgaty Downs, silicate	177.0	2.494	1.221	+0.18	0.69
Dalgaty Downs, metal	6.09	0.00946	-0.992	-3.50	0.23
Dalgaty Downs, WR	159.7	1.125	0.001	-1.85	0.46
Juvinas, WR	1,268	10.48	8.332	+9.84	0.56
Toluca, MC-ICPMS (Lyon), N = 1				-3.31	0.31
Toluca, MC-ICPMS (Harvard), N = 8				-3.16	0.15
Toluca, N-TIMS (Harvard), N = 7				-3.07	0.39

Data representation and analytical procedures. Here $f^{\text{Hf/W}} = ({}^{180}\text{Hf}/{}^{183}\text{W})_{\text{sample}} / ({}^{180}\text{Hf}/{}^{183}\text{W})_{\text{CHUR}} - 1$, where $({}^{180}\text{Hf}/{}^{183}\text{W})_{\text{CHUR}} = 2.836$ and the weight ratio $(\text{Hf}/\text{W})_{\text{CHUR}} = 1.123$ (refs 14, 15). Also $\epsilon_w = [({}^{182}\text{W}/{}^{183}\text{W})_{\text{sample}} / ({}^{182}\text{W}/{}^{183}\text{W})_{\text{standard}} - 1] \times 10^4$, using $({}^{182}\text{W}/{}^{183}\text{W})_{\text{standard}} = 1.85130 \pm 4$ (ref. 16) obtained for our terrestrial W standard (Harvard-W1). WR, 'whole rock' sample. The samples were all ~1–2 g, except for the CAI (~0.2 g). In selecting samples, we specifically avoided using un-equilibrated chondrites to construct internal isochrons, especially petrologic types between 3.0 and 3.9. This is because there is good evidence that metamorphism resulted in a net transfer of W from oxidized phases to metal before the complete consumption of reductants (carbides, organics, and so on), which may complicate the interpretation of the resulting Hf–W systematics. This process is considered complete for type 4 and higher²⁹. Therefore, an L4 chondrite, and an LL6 chondrite were selected for our work. The LL6 meteorite Dhurmsala was chosen for this study because its silicate fraction yielded a very high Hf/W ratio, which helps to define a precise isochron. Hf and W were separated at Harvard by anion exchange separation chemistry and their concentrations were determined precisely and accurately (to within $\pm 0.5\%$) using calibrated ¹⁸⁰Hf and ¹⁸⁴W isotopic tracers. Total procedural blanks for Hf and W are negligible (13–15 pg and <0.19 ng, respectively). For the BCR-1 standard we obtained Hf = 4.995 ± 24 p.p.b. and W = 412 ± 3 p.p.b., identical within error to accepted values for this standard, demonstrating the accuracy of our Hf–W ratios (analysed by multicollector inductively coupled plasma mass spectrometry (MC-ICP-MS) and N-TIMS at Harvard and MC-ICP-MS in Lyon). The W isotope data were corrected for instrumental mass fractionation using an exponential law with $({}^{184}\text{W}/{}^{183}\text{W}) = 2.14078$ and/or $({}^{186}\text{W}/{}^{183}\text{W}) = 1.98613 \pm 4$. The ϵ_w values in this table were obtained by MC-ICP-MS in Lyon, except for most of the Toluca measurements. The accuracy of our measurements of the ¹⁸²W/¹⁸³W ratio in meteorite samples relative to our terrestrial W standard (the ϵ_w value) was verified by comparing data for this meteorite (N = number of repeat measurements) obtained using three different mass spectrometers and two different techniques (MC-ICP-MS vs N-TIMS): (1) MC-ICP-MS using both the Micromass IsoProbe at Harvard as well as the VG Plasma 54 in Lyon, (2) N-TIMS using the Finnigan MAT262 at Harvard.

† Present address: Department of Earth & Planetary Sciences, University of Kobe, 1-1 Rokkoda-cho, Nada-Ku, Kobe, 657-8501 Japan.

inconsistent with the 3-Myr timescale implied by our present Hf–W data, as well as with evidence for now extinct ^{53}Mn (half-life $t_{1/2} = 3.7\text{ Myr}$, ref. 1) and ^{26}Al ($t_{1/2} = 0.7\text{ Myr}$, refs 2, 3) in eucrites. From our present Hf–W data and the available ^{53}Mn and ^{26}Al chronologies, we conclude that the eucrites formed $\sim 3\text{ Myr}$ after Allende's CAIs. The eucrite isochron in Fig. 1 exhibits a high initial ε_w of about -1 compared to the solar initial ε_w of about -3.5 . The mantle of the eucrite parent body (EPB; probably the asteroid Vesta) must have evolved as a high Hf/W reservoir in order to develop

radiogenic ^{182}W signatures. Core formation in the EPB resulted in a $f^{\text{La}/\text{W}} \approx f^{\text{Hf}/\text{W}} \approx 15$ for the EPB mantle²⁵. Thus, using the eucrite isochron⁶, the EPB mantle should have a present-day ε_w of $+17$. As 90–95% of W in the EPB is in its core, it follows by mass balance that the W isotope composition of the core is in the range 0.9 to 1.7 ε units below the present CHUR value of -2 . This is (within error) the same as the difference between present and initial CHUR of 1.5 ε units as obtained in our work, and also consistent with W isotope data for most iron meteorites and metals in chondrites^{8,15–19}. This requires that the time of eruption of these basalts on the surface of the EPB post-dates core formation in the EPB. Thus, the difficulties encountered in interpreting Hf–W systematics between eucrites and iron meteorites⁶ is resolved by using the chondrite data determined in the present study.

Using our present estimate of the CHUR values for the ^{182}Hf – ^{182}W system, we can explore the implications for the timing of accretion and core formation of the Earth. The difference between ε_w in the BSE and CHUR ($\Delta\varepsilon_w = \varepsilon_w(\text{BSE}) - \varepsilon_w(\text{CHUR})$) together with the $f^{\text{Hf}/\text{W}}$ of ~ 12 for the BSE¹⁵ provide the basis for such a calculation. The $\Delta\varepsilon_w$ value of the BSE is $+2$, and a plot of $\Delta\varepsilon_w$ versus the mean time of core formation is shown in Fig. 2. A two-stage model age for the BSE of 29 Myr since the formation of the

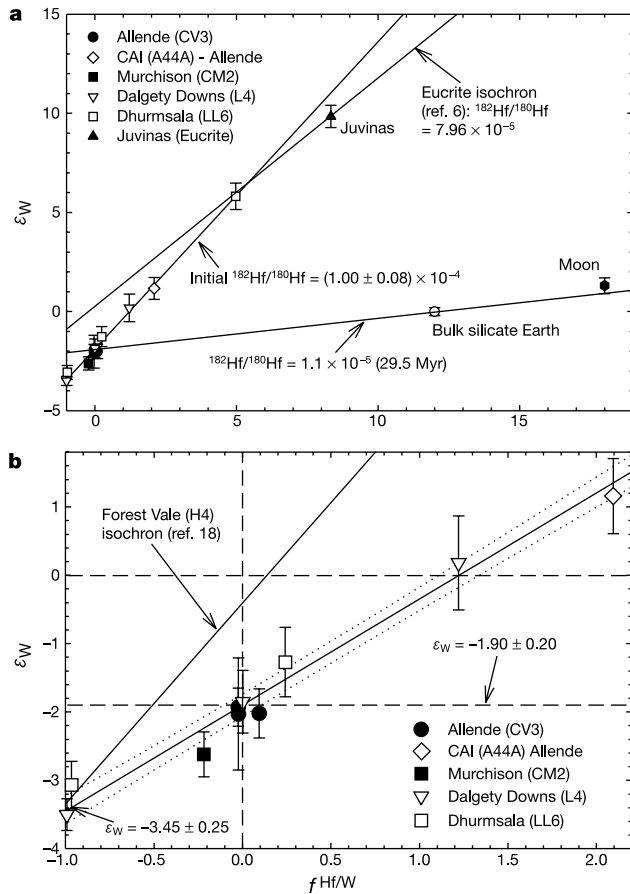


Figure 1 Hf–W systematics for the early Solar System. Shown is a plot of ε_w versus $^{180}\text{Hf}/^{183}\text{W}$ represented as $f^{\text{Hf}/\text{W}}$ (see Table 1 for definitions of ε_w and $f^{\text{Hf}/\text{W}}$). **a**, Data for metal and silicate fractions from ordinary chondrites Dalgety Downs (L4) and Dhurmsala (LL6), and from carbonaceous chondrites Allende and Murchison, define a good fossil isochron, identical within error of the individual isochrons for the two ordinary chondrites. Least-squares fitting of the data include the Allende and Murchison whole-rock data, but exclude the Allende CAI. Including or excluding the Murchison and Allende whole-rock data or the CAI data does not significantly change the slope or the intercept. Our Juvinas eucrite datum plots on the eucrite isochron⁶. The Moon, with a residual $\varepsilon_w = 1.3 \pm 0.4$ from ^{182}Hf decay²⁷ and $f^{\text{Hf}/\text{W}} = 18$ defined by the lunar La/W ratio²⁸, falls within error on the extension of the tie-line between the bulk chondrite (CHUR) and bulk silicate Earth (BSE) points. **b**, Magnified area for bulk chondrite data. Dotted curves show the 2σ error band. Our results are consistent with E-chondrite data¹⁹, the zircon data for the Simmern (H5) chondrite³⁰, Ste Marguerite (H4) and Richardton (H5) ordinary chondrites¹⁸, and inconsistent with published Allende and Murchison data^{7,8} and the published initial $^{182}\text{Hf}/^{180}\text{Hf}$ value of Forest Vale (H4)¹⁸. The position of our isochron relative to the Forest Vale isochron cannot be explained by late metamorphism. In that case the two isochrons would intersect at the bulk chondrite point ($f^{\text{Hf}/\text{W}} = 0$), which is not observed. Plausible reasons for discrepancy are: (1) uncorrected or improperly corrected interference; (2) contamination of the meteorite with terrestrial W either during or before chemical separation; and (3) incomplete dissolution, yielding a non-representative isotopic composition.

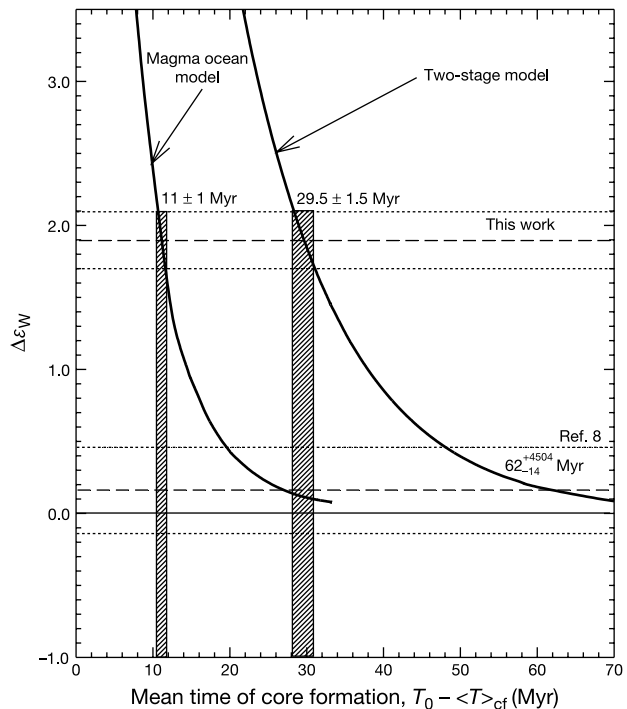


Figure 2 Models for timing of core formation in the Earth. Shown is the expected radiogenic $^{182}\text{W}/^{183}\text{W}$ value in the Earth relative to chondrites ($\Delta\varepsilon_w = [\varepsilon_w(\text{BSE}) - \varepsilon_w(\text{CHUR})]$) for a range of mean times of core formation (given by $T_0 - \langle T \rangle_{\text{cf}}$, where T_0 is the age of the Solar System and $\langle T \rangle_{\text{cf}}$ is the mean age of core formation; see ref. 14) in the Earth for two different models of core segregation: a two-stage model, and a magma ocean model. For the $\Delta\varepsilon_w$ value of $+1.9 \pm 0.20$ reported in this work, we obtain as shown a two-stage model age of $29.5 \pm 1.5\text{ Myr}$ and a mean time of core formation of $11 \pm 1.0\text{ Myr}$ (63% of core formed) for the continuous model (calculated as described in ref. 15). The vertical hatched bars represent the age uncertainties. We believe that the magma-ocean segregation model yields the most realistic estimate ($11 \pm 1.0\text{ Myr}$) for the mean time of core formation. For comparison, we show the previous results⁹ with a $\Delta\varepsilon_w$ value of $+0.17 \pm 0.29$. The timescale with the previous data⁹ is about a factor of $2^{+153}_{-0.3}$ longer (two-stage model age = $62^{+4504}_{-14}\text{ Myr}$) than the one obtained in this work. With the old data of ref. 8 the time of core formation is highly uncertain ($62^{+4504}_{-14}\text{ Myr}$), as the upper error limit corresponds to no radiogenic ^{182}W compared to chondrites in the silicate Earth.

Table 2 Comparison of new and old chronology

	New (this work)	Old (refs 7, 8, 18)
Initial $^{182}\text{Hf}/^{180}\text{Hf}$	$(1.00 \pm 0.08) \times 10^{-4}$	$(2.75 \pm 0.24) \times 10^{-4}$
Present bulk Solar System $^{182}\text{W}/^{183}\text{W}$ ratio	$\varepsilon_w = -1.90 \pm 0.20$	$\varepsilon_w = -0.17 \pm 0.29$
Earth core formation*	10–29 Myr	62^{+4504}_{-14} Myr
Moon*	29 Myr, and consistent with radiogenic ^{182}W on the Moon	>60 Myr, but inconsistent with radiogenic ^{182}W on the Moon
Vesta*	~3 Myr	16 Myr
Solar nebula gas dissipation	During planetary accretion	Before the main stage of planetary accretion
Mantle He and Ne isotope evidence	Consistent	Inconsistent

*Time given as Myr after the formation of the Solar System.

Solar System is shown as the intersection of the two-stage model line in Fig. 2 with $\Delta\varepsilon_w = +2$. Note that this corresponds to the 29-Myr fossil isochron for BSE in Fig. 1. This two-stage age has strict time significance only in the case where there is complete equilibration between the core and the silicate mantle at a single point in time, with no subsequent additions of material to the Earth. Thus, if Earth's accretion was terminated by a giant impact (giving rise to the Moon) and accompanied by complete metal–silicate equilibration, then the Hf/W system constrains this event to occur at 29 Myr after the beginning of the Solar System. However, as we have discussed earlier^{14–16,24}, continuous models of core formation should, in general, yield a more realistic time constraint. Here we consider a magma-ocean differentiation model¹⁵ with exponentially decreasing rates of accretion and with the rate of core formation limited by (and equal to) the accretion rate. In this model, all the metal segregated into the core is first equilibrated with the entire silicate mantle, as there is abundant evidence for an early magma ocean^{21,22,26}. This results in a mean time of core formation of 11 ± 1 Myr (63%) (and 24 Myr for 90%), substantially shorter than the two-stage model age. These results strongly suggest that the Earth's core formation took place earlier than stated in many recent publications^{7,8,20}.

Earlier reports⁹ required the Moon to have a substantially higher value of ε_w than the Earth. This made it extremely difficult to explain the Moon as being derived by a late (>50 Myr after the formation of the Solar System) giant impact at the end of Earth's accretion²⁰. It is now apparent that a significant portion of the observed ^{182}W excess in lunar samples was not the result of ^{182}Hf decay, but was instead the consequence of the $^{181}\text{Ta}(n,\gamma)^{182}\text{Ta}(\beta^-)^{182}\text{W}$ reaction induced by the bombardment of cosmic rays while these rocks were exposed on the surface of the Moon²⁷. The residual radiogenic ^{182}W signature from ^{182}Hf decay for the Moon is now $1.3 \pm 0.4 \varepsilon$ units higher than for the Earth²⁷. Therefore, W isotope data are consistent with formation of the Moon at ~25–30 Myr after the formation of the Solar System, as it plots within error of the BSE fossil isochron (see Fig. 1; here $f^{\text{Hf/W}}$ for the silicate Moon is estimated using the La/W ratio of the Moon²⁸). In fact, the absence of any differentiated silicate data below the chondrite–Earth–Moon line suggest that essentially the bulk of metal–silicate separation in the Solar System was completed in <30 Myr.

Astronomical observations place a severe constraint on the formation time of a gas-giant planet (possibly with a solid core of ten Earth masses) to within a few million years after central star formation and before complete dissipation of nebula gas^{4,5}. It is now well accepted from dynamical models that Mars-sized bodies will form within 0.1 Myr of the origin of the Solar System¹³. Dynamic accretion models favour the main growth stage (~60%) of the terrestrial planets (from Mars-sized to Earth-sized bodies) taking about 10–20 Myr, but the 'tail' of accretion may arguably continue for another 80–90 Myr (refs 11, 12). As shown here, our present W isotope data imply that the main growth stage is largely completed in about 10 Myr. Our data also resolve the inconsistency between Hf–W chronology and Mn–Cr and Al–Mg chronologies for meteorites derived from the Vesta asteroid, and support a very short timescale

(<3 Myr) for formation and melting of asteroid-sized bodies. We conclude that the Hf–W data reported here produce a consistent model for early Solar System chronology (see Table 2). □

Received 30 April; accepted 12 July 2002; doi:10.1038/nature00995.

- Lugmair, G. & Shukolyukov, A. Early solar system timescales according to ^{53}Mn – ^{53}Cr systematics. *Geochim. Cosmochim. Acta* **62**, 2863–2886 (1998).
- Srinivasan, G., Papanastassiou, D. A., Wasserburg, G. J., Bhandari, N. & Goswami, J. N. Re-examination of ^{26}Al – ^{26}Mg systematics in the Piplia Kalan eucrite. *Lunar Planet. Sci. XXXI*, A1795 (2000).
- Nyquist, L. E., Reese, Y., Wiesmann, H., Shih, C.-Y. & Takeda, H. Live ^{53}Mn and ^{26}Al in an unique cumulate eucrite with very calcic feldspar (An ~ 98). *Meteor. Planet. Sci. Suppl.* **36**, A151–A152 (2001).
- Briceño, C. *et al.* The CIDA-QUEST large-scale survey of Orion OB1: Evidence for rapid disk dissipation in a dispersed stellar population. *Science* **291**, 93–96 (2001).
- Bodenheimer, P. & Lin, D. N. C. Implications of extrasolar planets for understanding planet formation. *Annu. Rev. Earth Planet. Sci.* **30**, 113–148 (2002).
- Quitté, G., Birck, J.-L. & Allegre, C. J. ^{182}Hf – ^{182}W systematics in eucrites: the puzzle of iron segregation in the early solar system. *Earth Planet. Sci. Lett.* **184**, 83–94 (2000).
- Lee, D.-C. & Halliday, A. N. Hafnium–tungsten chronometry and the timing of terrestrial core formation. *Nature* **378**, 771–774 (1995).
- Lee, D.-C. & Halliday, A. N. Hf–W isotopic evidence for rapid accretion and differentiation in the early solar system. *Science* **274**, 1876–1879 (1996).
- Lee, D.-C., Halliday, A. N., Snyder, G. A. & Taylor, L. A. Age and origin of the Moon. *Science* **278**, 1098–1103 (1997).
- Halliday, A. N., Rehkämper, M., Lee, D.-C. & Yi, W. Early evolution of the Earth and Moon: new constraints from Hf–W isotope geochemistry. *Earth Planet. Sci. Lett.* **142**, 75–89 (1996).
- Wetherill, G. W. in *Origin of the Moon* (eds Hartmann, W. K., Phillips, R. J. & Taylor, G. J.) 519–550 (Lunar and Planetary Institute, Houston, 1986).
- Chambers, J. E. Making more terrestrial planets. *Icarus* **152**, 205–224 (2001).
- Kortenkamp, S. J., Wetherill, G. W. & Inaba, S. Runaway growth of planetary embryos facilitated by massive bodies in a protoplanetary disk. *Science* **293**, 1127–1129 (2001).
- Jacobsen, S. B. & Harper, C. L. Jr in *Earth Processes: Reading the Isotopic Code* Geophys. Monogr. 95 (eds Basu, A. & Hart, S.) 47–74 (American Geophysical Union, Washington DC, 1996).
- Harper, C. L. Jr & Jacobsen, S. B. Evidence for ^{182}Hf in the early Solar System and constraints on the timescale of terrestrial accretion and core formation. *Geochim. Cosmochim. Acta* **60**, 1131–1153 (1996).
- Jacobsen, S. B. & Yin, Q.-Z. W isotope variations and the time of formation of asteroidal cores and the Earth's core. *Lunar Planet. Sci. XXIX*, 1852–1853 (1998).
- Horan, M. F., Smoliar, M. I. & Walker, R. J. ^{182}W and ^{187}Re – ^{187}Os systematics of iron meteorites: chronology for melting, differentiation, and crystallization in asteroids. *Geochim. Cosmochim. Acta* **62**, 545–554 (1998).
- Lee, D.-C. & Halliday, A. N. Hf–W internal isochrons for ordinary chondrites and the initial $^{182}\text{Hf}/^{180}\text{Hf}$ of the solar system. *Chem. Geol.* **169**, 35–43 (2000).
- Lee, D.-C. & Halliday, A. N. Accretion of primitive planetesimals: Hf–W isotopic evidence from enstatite chondrites. *Science* **288**, 1629–1631 (2000).
- Halliday, A. N., Lee, D.-C. & Jacobsen, S. B. in *Origin of the Earth and Moon* (eds Canup, R. M. & Righter, K.) 45–62 (Univ. Arizona Press, Tucson, 2000).
- Harper, C. & Jacobsen, S. B. Noble gases and Earth's accretion. *Science* **273**, 1814–1818 (1996).
- Porcelli, D., Woolum, D. & Cassen, P. Deep Earth rare gases: initial inventories, capture from the solar nebula, and losses during Moon formation. *Earth Planet. Sci. Lett.* **193**, 237–251 (2001).
- Cameron, A. G. W. From interstellar gas to the Earth–Moon system. *Meteorit. Planet. Sci.* **36**, 9–22 (2001).
- Yin, Q.-Z. *et al.* New Hf–W data that are consistent with Mn–Cr chronology: implications for early solar system evolution. *Lunar Planet. Sci. XXXIII*, A1700 (2002).
- Palme, H. & Rammensee, W. The significance of W in planetary differentiation processes: Evidence from new data on eucrites. *Proc. Lunar Planet. Sci. Conf. B* **12**, 949–964 (1981).
- Drake, M. J. & Righter, K. Determining the composition of the Earth. *Nature* **416**, 39–44 (2002).
- Lee, D.-C., Halliday, A. N., Leya, I., Wieler, R. & Wiechert, U. Cosmogenic tungsten and the origin and earliest differentiation of the Moon. *Earth Planet. Sci. Lett.* **198**, 267–274 (2002).
- Rammensee, W. & Wänke, H. On the partition coefficient of tungsten between metal and silicate and its bearing on the origin of the moon. *Proc. Lunar Sci. Conf.* **8**, 399–409 (1977).
- Humayun, M. & Campbell, A. J. The duration of ordinary chondrite metamorphism inferred from tungsten microdistribution in metal. *Earth Planet. Sci. Lett.* **198**, 225–243 (2002).
- Ireland, T. R., Kirby, H., Bukovanska, M. & Wlotzka, F. Hf–W systematics of meteoritic zircons, revisited. *Lunar Planet. Sci. XXXI*, A1540 (2000).

Acknowledgements

We thank the Smithsonian Institution, Harvard Mineralogical Museum, U. Marvin and H. Palme for providing the samples, and A. N. Halliday and D.-C. Lee for comments on this paper. This work was supported by NASA's Cosmochemistry and Origin of Solar System programmes and the National Science Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Q.Z.Y. (e-mail: yin@fas.harvard.edu).

Rapid accretion and early core formation on asteroids and the terrestrial planets from Hf–W chronometry

T. Kleine*, C. Münker*, K. Mezger* & H. Palme†

* Institut für Mineralogie, Universität Münster, Corrensstrasse 24, D-48149 Münster, Germany

† Institut für Mineralogie und Geochemie, Universität zu Köln, Zùlpicherstrasse 49b, D-50674 Köln, Germany

The timescales and mechanisms for the formation and chemical differentiation of the planets can be quantified using the radioactive decay of short-lived isotopes^{1–10}. Of these, the ¹⁸²Hf-to-¹⁸²W decay is ideally suited for dating core formation in planetary bodies^{1–5}. In an earlier study, the W isotope composition¹ of the Earth's mantle was used to infer that core formation was late¹ (≥ 60 million years after the beginning of the Solar System) and that accretion was a protracted process^{11,12}. The correct interpretation of Hf–W data depends, however, on accurate knowledge of the initial abundance of ¹⁸²Hf in the Solar System and the W isotope composition of chondritic meteorites. Here we report Hf–W data for carbonaceous and H chondrite meteorites that lead to timescales of accretion and core formation significantly different from those calculated previously^{1,3,5,11,12}. The revised ages for Vesta, Mars and Earth indicate rapid accretion, and show that the timescale for core formation decreases with decreasing size of the planet. We conclude that core formation in the terrestrial planets and the formation of the Moon must have occurred during the first ~ 30 million years of the life of the Solar System.

Chemically and mineralogically, carbonaceous chondrites represent some of the most primitive material in the Solar System. The chemical composition of type 1 (CI) carbonaceous chondrites is, except for extremely volatile elements (H, C, N, O, rare gases), identical to that of the bulk Solar System. Other types of carbonaceous chondrites are fractionated relative to CI chondrites, but refractory elements occur in chondritic (solar) proportions in all types. Hence, all carbonaceous chondrites should have the same ratio of the two refractory elements Hf and W, and therefore a uniform W isotope composition. Variations in the Hf/W ratio among different groups of carbonaceous chondrites of up to $\sim 30\%$ would result in W isotope variations indistinguishable at the currently obtained analytical resolution. We analysed the W isotope compositions of seven carbonaceous chondrites—Orgueil (CI), Allende (CV), Murray (CM), Murchison (CM), Cold Bokkeveld (CM), Nogoya (CM) and Karoonda (CK)—using multicollector inductively coupled plasma mass spectrometry (MC-ICP-MS),

and found them to be identical within analytical error. The weighted average of the ¹⁸²W/¹⁸⁴W ratios for these samples differs by $-1.9 \pm 0.2 \epsilon$ units from the terrestrial standard (Table 1, Fig. 1), but agrees with Hf–W data for enstatite chondrites¹³.

The most reliable method for determining the initial abundance of ¹⁸²Hf is by measuring internal Hf–W isochrons of samples with independently known age. The H chondrites Ste Marguerite and Forest Vale were selected because their phosphates have been dated previously by the high-precision Pb–Pb method. The preservation of U–Pb phosphate ages of 4.5627 ± 0.0006 Gyr for Ste Marguerite¹⁴ and 4.5609 ± 0.0007 Gyr for Forest Vale¹⁴ precludes late disturbance or alteration of these meteorites. A four-point metal-silicate isochron for Ste Marguerite (Fig. 2a) defines an initial ¹⁸²Hf/¹⁸⁰Hf ratio of $(0.85 \pm 0.05) \times 10^{-4}$ at the time of closure of the Hf–W system. Likewise, a four-point isochron for Forest Vale gives an initial ¹⁸²Hf/¹⁸⁰Hf ratio of $(1.0 \pm 0.5) \times 10^{-4}$ (Fig. 2b). These two isochrons pass through the newly defined chondritic ϵ_w at the chondritic Hf/W ratio of ~ 1.1 .

Back calculation of these initial values to the Pb–Pb age of Ca–Al-rich inclusions (CAI; 4.566 ± 0.002 Gyr, ref. 15), which is commonly used as reference for condensation of the first solid matter in the Solar System, yields a ¹⁸²Hf/¹⁸⁰Hf ratio of $(1.09 \pm 0.09) \times 10^{-4}$ for Ste Marguerite and $(1.5 \pm 0.7) \times 10^{-4}$ for Forest Vale. These values define the initial abundance of ¹⁸²Hf in the Solar System, provided that the ¹⁸²Hf–¹⁸²W system in metal and silicate closed at approximately the same time as the U–Pb system in phosphates. This is likely, as these H chondrites do not show evidence for later thermal overprint. The more precise value of $(1.09 \pm 0.09) \times 10^{-4}$ for Ste Marguerite is our preferred approximation for the initial ¹⁸²Hf/¹⁸⁰Hf ratio of the Solar System. This precisely defined value is in agreement with previous estimates obtained from internal chondrite isochrons¹⁶, the comparison of W isotopes in iron meteorites and chondrites^{16,30}, and the W isotope compositions of

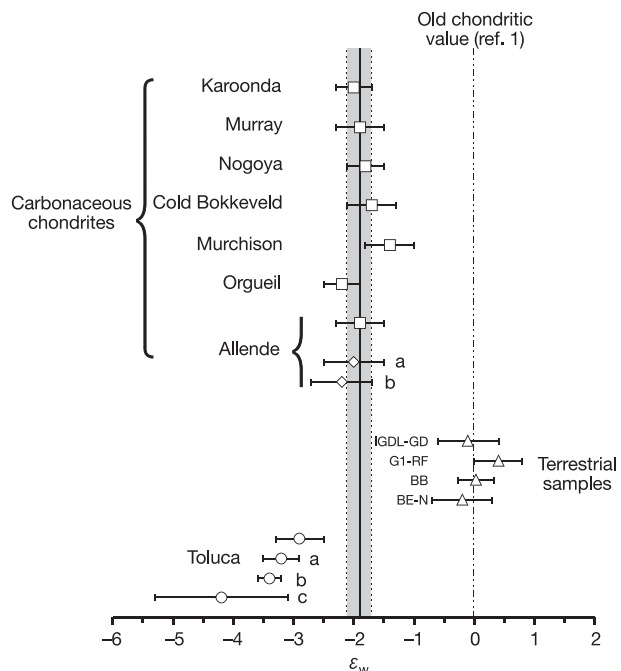


Figure 1 ϵ_w values of carbonaceous chondrites compared with those of the Toluca iron meteorite and terrestrial samples analysed in this study. The values for Toluca, Allende, G1-RF and IGDL-GD are the weighted averages of four or more independent analyses. Also included are data from ref. 16 (indicated by a), ref. 30 (b), and ref. 2 (c). For the definition of ϵ_w see Table 1. The vertical shaded bar refers to the uncertainty in the W isotope composition of chondrites. Terrestrial samples include IGDL-GD (greywacke), G1-RF (granite) and BB and BE-N (basalts).

meteoritic zircons¹⁷. This newly determined value is significantly lower than the previous estimate¹⁸ of $(2.75 \pm 0.24) \times 10^{-4}$.

The revision of both the chondritic W isotope composition and the initial $^{182}\text{Hf}/^{180}\text{Hf}$ has fundamental implications for the Hf–W chronometry of planetary core formation. Assuming that core formation occurred as a single-stage event, the time of metal–silicate separation can be calculated using the differences in the W isotope composition and Hf/W between a silicate or metal reservoir and chondrites¹.

All known classes of iron meteorites exhibit ε_{W} anomalies ranging from -5 to -3 relative to the terrestrial value⁴, indicating that core formation in iron meteorite parent bodies took place within 10 Myr of the beginning of the Solar System (defined here as the condensation of the oldest dated phases in the Solar System). This constraint is consistent with the timescale deduced from ^{107}Pd – ^{107}Ag systematics in iron meteorites^{9,10}. Owing to the low Solar System initial $^{182}\text{Hf}/^{180}\text{Hf}$ of $(1.09 \pm 0.09) \times 10^{-4}$ and the low Hf/W in metals, variations in the W isotope composition of iron meteorites and thus time differences for metal differentiation cannot be resolved with current analytical techniques. However, precise timescales for core formation can be obtained from the silicate portions of planetary bodies (Vesta⁵, Mars³, Earth and Moon¹⁹), which have high Hf/W. Published Hf–W data for eucrites, which are widely considered to be from the asteroid Vesta²⁰, define a whole-rock isochron corresponding to an initial $^{182}\text{Hf}/^{180}\text{Hf}$ of $(7.96 \pm 0.34) \times 10^{-5}$ (ref. 5). Using our new parameters, this yields a Hf–W age for silicate differentiation on Vesta of 4.2 ± 1.3 Myr after the beginning of the Solar System. With the old Hf–W parameters, a value of 16.1 ± 1.3 Myr was reported⁵; this young age was difficult to reconcile with the general view that decay of ^{26}Al , which was only present in significant amounts during the first ~ 5 Myr of the Solar System, was the main heat source for internal differentiation of planetesimals. In fact, the former presence of ^{26}Al has been detected in the Piplia Kalan eucrite⁸. Our revised age of 4.2 ± 1.3 Myr after the beginning of the Solar System is in notable agreement with timescales deduced from the ^{53}Mn – ^{53}Cr (ref. 6), ^{60}Fe – ^{60}Ni (ref. 7) and ^{26}Al – ^{26}Mg chronometers⁸, which all suggest mantle–crust differentiation on Vesta within the first ~ 4 Myr of the Solar System.

To obtain direct age constraints for core formation on Vesta, the W isotope composition and the Hf/W of the bulk mantle of Vesta

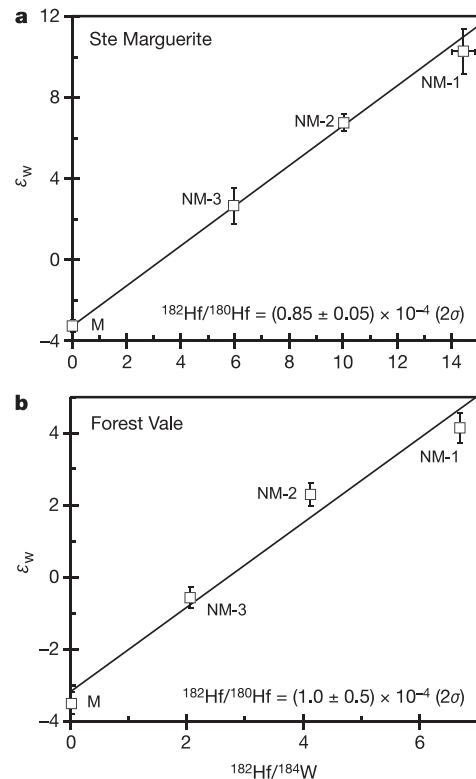


Figure 2 ε_{W} versus $^{180}\text{Hf}/^{184}\text{W}$ for different fractions of the H chondrites Ste Marguerite (a) and Forest Vale (b). NM-1, NM-2 and NM-3 refer to different nonmagnetic fractions, M is the magnetic fraction. We interpret the positive correlation of ε_{W} with $^{180}\text{Hf}/^{184}\text{W}$ as an internal Hf–W isochron whose slope corresponds to the initial $^{182}\text{Hf}/^{180}\text{Hf}$ ratio at the time of closure of the Hf–W system.

need to be known. Using a Hf/W of 34.9 (calculated from W/La and La/Hf; refs 21, 22) for the bulk mantle of Vesta, we deduce a W isotope composition of $+35.8 \varepsilon_{\text{W}}$ from the eucrite Hf–W whole-rock isochron, yielding a W model age for core formation of 3.8 ± 1.3 Myr since the beginning of the Solar System. This age is

Table 1 W isotope data for carbonaceous and H chondrites

Sample	Hf (p.p.b.)	W (p.p.b.)	$^{180}\text{Hf}/^{184}\text{W}$ $\pm 2\sigma$	$^{182}\text{W}/^{184}\text{W}$ $\pm 2\sigma$	$^{183}\text{W}/^{184}\text{W}$ $\pm 2\sigma$	ε_{W} $\pm 2\sigma$
Carbonaceous chondrites						
Allende				0.864532 ± 35	0.467052 ± 23	-1.9 ± 0.4
Orgueil				0.864506 ± 26	0.467122 ± 14	-2.2 ± 0.3
Murchison				0.864575 ± 35	0.467052 ± 19	-1.4 ± 0.4
Cold Bokkeveld				0.864549 ± 35	0.467066 ± 19	-1.7 ± 0.4
Nogoya				0.864540 ± 26	0.467052 ± 14	-1.8 ± 0.3
Murray				0.864532 ± 35	0.467047 ± 19	-1.9 ± 0.4
Karoonda				0.864523 ± 26	0.467038 ± 14	-2.0 ± 0.3
Weighted average				0.864533 ± 20	0.467050 ± 10	-1.9 ± 0.2
H chondrites						
Ste Marguerite						
M	11.6	856.2	0.0155 ± 2	0.864412 ± 26	0.467067 ± 14	-3.3 ± 0.3
NM-1	204.9	16.2	14.4 ± 4	0.865583 ± 95	0.467068 ± 50	10.3 ± 1.1
NM-2	148.8	17.0	10.0 ± 1	0.865279 ± 35	0.467049 ± 14	6.7 ± 0.4
NM-3	183.3	35.3	5.9 ± 2	0.864925 ± 78	0.467010 ± 34	2.7 ± 0.9
Forest Vale						
M	8.22	664.5	0.0142 ± 1	0.864394 ± 26	0.467060 ± 12	-3.5 ± 0.3
NM-1	179.0	30.7	6.68 ± 6	0.865054 ± 36	0.467088 ± 16	4.1 ± 0.4
NM-2	148.6	41.2	4.13 ± 3	0.864896 ± 28	0.467060 ± 14	2.3 ± 0.3
NM-3	170.0	95.0	2.05 ± 1	0.864648 ± 26	0.467080 ± 13	-0.6 ± 0.3

The W isotope composition of a sample is generally reported in ε_{W} units relative to a terrestrial standard, given by $\varepsilon_{\text{W}} = [(^{182}\text{W}/^{184}\text{W})_{\text{sample}} / (^{182}\text{W}/^{184}\text{W})_{\text{standard}} - 1] \times 10^4$. Uncertainties refer to the last significant digits. The accuracy of the measured W isotope ratios is monitored by the analysis of $^{183}\text{W}/^{184}\text{W}$. For all samples except Orgueil, the $^{183}\text{W}/^{184}\text{W}$ ratios agree with the terrestrial standard. The higher $^{183}\text{W}/^{184}\text{W}$ for Orgueil is caused by an organic peak on mass 183 that for all other samples was successfully removed by treatment with $\text{HNO}_3\text{--H}_2\text{O}_2$. For the H chondrites, M and NM denote magnetic and nonmagnetic fractions, respectively.

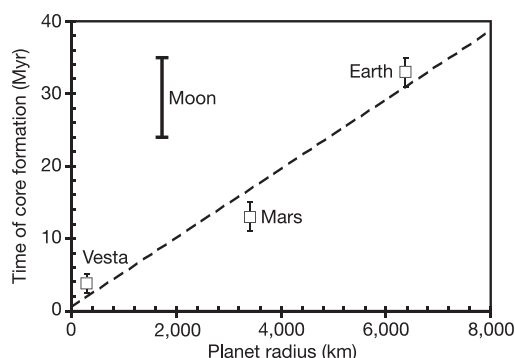


Figure 3 Time of core formation in Myr after CAI condensation for Vesta, Mars, Earth and Moon versus planet radius as deduced from Hf–W systematics. For the Moon, the two data points refer to the endmember model ages. The Moon plots distinctly to the left of the correlation line defined by Vesta, Mars and Earth, suggesting a different formation process.

indistinguishable from the silicate differentiation age obtained from the eucrite whole-rock isochron. ^{182}Hf – ^{182}W systematics therefore show that both core formation and silicate differentiation on Vesta occurred within the first ~5 Myr of the Solar System.

Martian meteorites display W isotope compositions ranging from $+0.3\epsilon_w$ to $+2.9\epsilon_w$ (ref. 3). Correlated ^{182}W and ^{142}Nd anomalies on Mars indicate that the observed spread in ϵ_w is caused by internal silicate differentiation in the martian mantle³ because Sm and Nd are both lithophile elements that do not fractionate during core formation. As martian meteorites with no excess ^{142}Nd have $^{182}\text{W}/^{184}\text{W}$ ratios that are identical to the earlier chondritic value, core formation and silicate differentiation on Mars were interpreted as contemporaneous processes that occurred within the first 30 Myr of the Solar System (ref. 3). However, with the revised chondritic W isotope composition, even the least radiogenic martian meteorites display a ^{182}W excess of ~2.2 ϵ units relative to the new chondritic value. As the $^{142}\text{Nd}/^{144}\text{Nd}$ ratio in these meteorites is chondritic, the increased $^{182}\text{W}/^{184}\text{W}$ can be uniquely attributed to Hf–W fractionation that is caused by core formation, suggesting a decoupling of core formation and silicate differentiation on Mars. Assuming a Hf/W of 5.1 for the primitive martian mantle^{22,23}, the new W model age for core formation on Mars is 13 ± 2 Myr after the beginning of the Solar System.

Revision of the ^{182}Hf – ^{182}W parameters has far-reaching implications for the timing of terrestrial core formation and the origin of the Earth–Moon system. Most importantly, the bulk silicate Earth (BSE) has a clearly resolvable ^{182}W excess of 1.9 ϵ units relative to chondrites, and does not have a chondritic W isotope composition as previously suggested¹. Using a Hf/W ratio of 17.7 for the BSE^{24,25}, core formation on Earth occurred 33 ± 2 Myr after the beginning of the Solar System. The W isotope composition of the lunar mantle ranges from $+1.3 \pm 0.4$ to the terrestrial value of $0\epsilon_w$ (ref. 19). Assuming that the Hf/W ratios of the lunar mantle and BSE are identical¹, W model ages for the Moon are between 26 ± 2 and 33 ± 2 Myr after the beginning of the Solar System, and thus substantially older than previously suggested¹⁹.

The revised Hf–W model ages for core formation indicate rapid accretion and early core formation on planetary bodies, and an early origin of the Moon. These timescales are consistent with the rates suggested by Wetherill-type accretion models^{26–28}. The fact that the time of core formation increases with planet size suggests a more protracted growth history for the larger planets (Fig. 3). The Moon does not fit this simple relationship, indicating that it formed by a different process—such as a giant impact during Earth's early accretion. □

Methods

Sample preparation

Samples were cleaned in an ultrasonic bath, washed in 0.05 M HNO_3 for a few minutes, and powdered in an agate mortar. Metal–silicate separation for the H chondrites was performed using a hand magnet. Silicate dust adhering to the metal grains was removed by repeated crushing of the magnetic fractions in ethanol. Some non-magnetic fractions were further purified by alternately grinding the powder in an agate mortar and removing the remaining metal grains with a magnet. Carbonaceous chondrites and the non-magnetic fractions from H chondrites were dissolved in Savillex vials at 180°C on a hotplate using HF – HNO_3 – HClO_4 (5:3:2). Metal separates from H chondrites were dissolved in 5:1 HNO_3 – HF at 120°C on a hotplate. After digestion, the samples were dried and redissolved in HNO_3 – H_2O_2 to remove organic compounds. After drying down, the samples were completely dissolved in 6 M HCl –0.06 M HF . A 10–20% aliquot of the sample was spiked with a mixed ^{180}Hf – ^{183}W tracer that was calibrated against pure Hf and W metals. The use of high-pressure bombs for sample digestion was specifically avoided because Savillex vials contain significant W (up to several nanograms) that is released at high pressure. Furthermore, W is adsorbed by Savillex Teflon, thus requiring thorough cleaning of the beakers. For this reason, the W blank of every vial was checked in a blank digestion procedure before each sample digestion. This procedure was found to be absolutely necessary to achieve low and reproducible blanks. Separation of Hf and W from the matrix was performed by anion exchange (AG-1X8 resin) in HCl – HF following procedures modified from refs 4 and 29. Hafnium was eluted using 9 M HCl –0.01 M HF , and W was collected in 7 M HCl –1 M HF . Total procedural blanks were ~250 pg for W isotope composition measurements, and ~50 pg W and ~10 pg Hf for the concentration measurements.

Isotope measurements

All measurements were made using the IsoProbe MC-ICP-MS at Münster. Tungsten isotope measurements were normalized to $^{186}\text{W}/^{184}\text{W} = 0.92767$ using the exponential law. Repeated measurements of a ALFA AESAR standard metal during the course of this study yielded $^{182}\text{W}/^{184}\text{W} = 0.864696 \pm 0.000009$ (2σ of the mean) and $^{183}\text{W}/^{184}\text{W} = 0.467052 \pm 0.000006$ (2σ of the mean). The accuracy and external reproducibility of the W isotope measurements ($\pm 0.5\epsilon$, 2σ) were determined by repeated measurements of the Toluca iron meteorite and the Allende carbonaceous chondrite (Fig. 1). The robustness of our method is furthermore documented by several independent measurements (separate digestions) of different terrestrial samples (Fig. 1). Small isobaric Os interferences on masses 186 and 184 were corrected by monitoring ^{188}Os . To assess possible interferences on ^{186}W and ^{184}W , measured $^{182}\text{W}/^{184}\text{W}$ values were also normalized to $^{186}\text{W}/^{183}\text{W}$ and $^{183}\text{W}/^{184}\text{W}$ using the exponential law. For the chondrites identical results of -1.9 ± 0.2 and $-1.8 \pm 0.4\epsilon$ units are obtained when using $^{186}\text{W}/^{183}\text{W}$ and $^{183}\text{W}/^{184}\text{W}$, respectively, for mass bias correction.

Received 30 May; accepted 17 July 2002; doi:10.1038/nature00982.

- Lee, D. C. & Halliday, A. N. Hafnium–tungsten chronometry and the timing of terrestrial core formation. *Nature* **378**, 771–774 (1995).
- Harper, C. L. & Jacobsen, S. B. Evidence for ^{182}Hf in the early solar system and constraints on the timescale of terrestrial accretion and core formation. *Geochim. Cosmochim. Acta* **60**, 1131–1153 (1996).
- Lee, D. C. & Halliday, A. N. Core formation on Mars and differentiated asteroids. *Nature* **388**, 854–857 (1997).
- Horan, M. F., Smoliar, M. I. & Walker, R. J. ^{182}W and ^{187}Re – ^{187}Os systematics of iron meteorites: Chronology for melting, differentiation, and crystallization in asteroids. *Geochim. Cosmochim. Acta* **62**, 545–554 (1998).
- Quitté, G., Birc, J. L. & Allège, C. J. ^{182}Hf – ^{180}Hf systematics in eucrites: the puzzle of iron segregation in the early solar system. *Earth Planet. Sci. Lett.* **184**, 83–94 (2000).
- Lugmair, G. W. & Shukolyukov, A. Early solar system timescales according to ^{53}Mn – ^{53}Cr systematics. *Geochim. Cosmochim. Acta* **62**, 2863–2886 (1998).
- Shukolyukov, A. & Lugmair, G. W. ^{60}Fe in eucrites. *Earth Planet. Sci. Lett.* **119**, 159–166 (1993).
- Srinivasan, G., Goswami, J. & Bhandari, N. ^{26}Al in eucrite Piplia Kalan: plausible heat source and formation chronology. *Science* **284**, 1348–1350 (1999).
- Chen, J. H. & Wasserburg, G. J. The isotopic composition of Ag in meteorites and the presence of ^{107}Pd in protoplanets. *Geochim. Cosmochim. Acta* **54**, 1729–1743 (1990).
- Carlson, R. W. & Hauri, E. H. Extending the ^{107}Pd – ^{107}Ag chronometer to low Pd/Ag meteorites with multicollector plasma-ionization mass spectrometry. *Geochim. Cosmochim. Acta* **65**, 1839–1848 (2001).
- Halliday, A. N. & Lee, D. C. Tungsten isotopes and the early development of the earth and the moon. *Geochim. Cosmochim. Acta* **63**, 4157–4179 (1999).
- Shearer, C. K. & Newsom, H. E. W–Hf isotope abundances and the early origin and evolution of the Earth–Moon system. *Geochim. Cosmochim. Acta* **64**, 3599–3613 (2000).
- Lee, D. C. & Halliday, A. N. Accretion of primitive planetesimals: Hf–W isotopic evidence from enstatite chondrites. *Science* **288**, 1629–1631 (2000).
- Göpel, C., Manhès, G. & Allège, C. J. U–Pb systematics of phosphates from equilibrated ordinary chondrites. *Earth Planet. Sci. Lett.* **121**, 153–171 (1994).
- Allège, C. J., Manhès, G. & Göpel, C. The age of the Earth. *Geochim. Cosmochim. Acta* **59**, 1445–1456 (1995).
- Yin, Q. *et al.* New Hf–W data that are consistent with Mn–Cr chronology: implications for early solar system evolution. *Lunar Planet. Sci.* **XXXIII**, 1700 (2002).
- Ireland, T. R., Kirby, H., Bukovanska, M. & Wlotzka, F. Hf–W systematics of meteoritic zircons, revisited. *Lunar Planet. Sci.* **XXXI**, 1540 (2000).
- Lee, D. C. & Halliday, A. N. Hf–W internal isochrons for ordinary chondrites and the initial $^{182}\text{Hf}/^{180}\text{Hf}$ of the solar system. *Chem. Geol.* **169**, 35–43 (2000).
- Lee, D. C. & Halliday, A. N. Cosmogenic tungsten and the origin and earliest differentiation of the Moon. *Earth Planet. Sci. Lett.* **198**, 267–274 (2002).

20. Binzel, R. P. & Xu, S. Chips off of asteroid 4 Vesta: evidence for the parent body of basaltic achondrite meteorites. *Science* **260**, 186–191 (1993).
21. Palme, H. & Rammensee, W. The significance of W in planetary differentiation processes: Evidence from new data on eucrites. *Lunar Planet. Sci. XXII*, 949–964 (1981).
22. Palme, H. & Beer, H. in *Landolt-Börnstein Group VI, Astronomy and Astrophysics Vol. 3a, Instruments; Methods; Solar System* (ed. Voigt, H. H.) 196–221 (Springer, Berlin, 1993).
23. Kong, P., Ebihara, M. & Palme, H. Siderophile elements in Martian meteorites and implications for core formation in Mars. *Geochim. Cosmochim. Acta* **63**, 1865–1875 (1999).
24. Newsom, H. E. *et al.* The depletion of tungsten in the bulk silicate earth: Constraints on core formation. *Geochim. Cosmochim. Acta* **60**, 1155–1169 (1996).
25. McDonough, W. F. & Sun, S.-S. The composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).
26. Wetherill, G. W. Formation of the Earth. *Annu. Rev. Earth Planet. Sci.* **18**, 205–256 (1990).
27. Canup, R. M. & Agnor, C. in *Origin of Earth and Moon* (eds Canup, R. M. & Righter, K.) 1839–1848 (Cambridge Univ. Press, Cambridge, 2001).
28. Chambers, J. E. Making more terrestrial planets. *Icarus* **152**, 205–224 (2001).
29. Münker, C., Weyer, S., Scherer, E. & Mezger, K. Separation of high field strength elements (Nb, Ta, Zr, Hf) and Lu from rock samples for MC-ICPMS measurements. *Geochem. Geophys. Geosyst.* **2**, 2001GC000183 (2001).
30. Schoenberg, R., Kamber, B. S., Collerson, K. D. & Eugster, O. New W-isotope evidence for rapid terrestrial accretion and very early core formation. *Geochim. Cosmochim. Acta* (in the press).

Acknowledgements

We thank the Muséum National d'Histoire Naturelle (Paris) for providing the meteorites Ste Marguerite and Forest Vale, the MPI Mainz for the carbonaceous chondrites, E. Scherer for discussions, and M. Drake, A. Halliday and D. Papanastassiou for comments and suggestions. This work was supported by the Deutsche Forschungsgemeinschaft as part of the research priority programme 'Mars and the terrestrial planets'.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.K. (e-mail: tkleine@nwz.uni-muenster.de).

Metal–insulator transition in chains with correlated disorder

Pedro Carpena*†, Pedro Bernal-Galván*†, Plamen Ch. Ivanov† & H. Eugene Stanley†

* Departamento de Física Aplicada II, ETSI de Telecomunicación, Universidad de Málaga, 29071 Málaga, Spain

† Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA

According to Bloch's theorem, electronic wavefunctions in perfectly ordered crystals are extended, which implies that the probability of finding an electron is the same over the entire crystal¹. Such extended states can lead to metallic behaviour. But when disorder is introduced in the crystal, electron states can become localized, and the system can undergo a metal–insulator transition (also known as an Anderson transition)^{2–4}. Here we theoretically investigate the effect on the physical properties of the electron wavefunctions of introducing long-range correlations in the disorder in one-dimensional binary solids, and find a correlation-induced metal–insulator transition. We perform numerical simulations using a one-dimensional tight-binding model, and find a threshold value for the exponent characterizing the long-range correlations of the system. Above this threshold, and in the thermodynamic limit, the system behaves as a conductor within a broad energy band; below threshold, the system behaves as an insulator. We discuss the possible relevance of this result for electronic transport in DNA, which displays long-range correlations^{5,6} and has recently been reported to be a one-dimensional disordered conductor^{7–10}.

The tight-binding model of a solid is characterized by a hamiltonian in which one orbital (single electron) and a single energy ε_i

are assigned to each lattice site i

$$H = \sum_i \varepsilon_i |i\rangle\langle i| + \sum_{\langle i,j \rangle} t |i\rangle\langle j| \quad (1)$$

where t is the electronic overlap (hopping term) between the wavefunctions of electrons centred at two neighbouring sites i, j . For a perfectly ordered crystal, all site energies ε_i have the same value or follow a periodic pattern. For disordered solids, ε_i can be randomly chosen from a certain probability distribution—for example, a uniform distribution or a gaussian. The probability distribution generating ε_i is characterized by a parameter W which quantifies the spread of the distribution, and thus the degree of disorder. In the case of a uniform distribution W is the distribution width, whereas in the case of a gaussian distribution W is the standard deviation. Models for one-dimensional (1D) disordered solids traditionally consider only two parameters of interest, namely the interaction term between nearest neighbours, t , and the disorder of the system, W . Thus the controlling parameter is the ratio W/t . Typically, t is fixed (for example, $t = 1$), thus fixing the energy scale in the system, and W is changed to study the effects produced by different levels of disorder.

In the limit $W \rightarrow 0$, where the coupling energy t between neighbours dominates, a perfect lattice is recovered, and Bloch's theorem applies: there are extended electron states, and the system can conduct¹. This is valid for finite system size only, because localization theory asserts that in the limit of large systems all electron states are localized in the 1D case³. For sufficiently large W , where the disorder dominates, the wavefunctions are strongly localized. This localization phenomenon depends on the system dimension. For three-dimensional (3D) systems at zero temperature, $T = 0$, there is a critical value W_c for which a metal–insulator transition occurs^{11,12}. For a uniform distribution, $W_c = 16.5$. For $W < W_c$, despite some degree of disorder, the electron wavefunctions are extended, and the system behaves as a metal, whereas for $W > W_c$, the wavefunctions become localized, and the system behaves as an insulator. For two-dimensional (2D) and 1D disordered systems even infinitesimal disorder produces localized states³, so at $T = 0$ the system behaves as an insulator in the thermodynamic limit, although for some particular cases in two dimensions it is possible to find a metal–insulator transition^{13–15}.

We consider the case of a 1D system, and show that randomly chosen but long-range correlated $\{\varepsilon_i\}$ can lead to extended wavefunctions and thus to conductivity, in contrast to the expectation—based on the assumption of uncorrelated disorder—that no extended states can be found in 1D disordered systems at $T = 0$ (ref. 1). Traditionally, localization of the electron wavefunction in one dimension is considered to be unaffected by the form and width of the distribution from which the $\{\varepsilon_i\}$ are chosen. Although all proofs establishing localization in one dimension are model-dependent, exponential localization of all eigenstates in one dimension is believed to occur at $T = 0$ (refs 16–18). In this case, electron wavefunctions are of the form $\Psi(x) = f(x) \exp(-|x - x_0|/\lambda)$, where $f(x)$ is a random function which depends on the particular realization of the disordered chain, and λ is the localization length, which is a measure of the size of the wavefunction.

The simplest 1D model exhibiting Anderson localization is the random binary alloy¹⁹, where there are two types of atoms, A and B, and two different diagonal energies, ε_A and ε_B . To build the series of diagonal energies of the hamiltonian (equation (1)), ε_A and ε_B are assigned at random to each lattice site with probabilities p and $1 - p$, respectively. For the random binary alloy, the corresponding wavefunctions are localized and the system behaves as an insulator (Fig. 1a).

We now show that introducing correlations in the disorder can markedly change the physics. The first attempt to introduce correlations into the random binary alloy model was the random dimer model^{20,21}. In this model, short-range correlations are introduced by

assigning a dimer AA with probability p , and a monomer B with probability $1 - p$. The main property of the random dimer model is that there exists a particular energy (the resonant energy of the dimer) for which there is perfect electron transmission through the system. Wavefunctions corresponding to energies close to this resonant energy behave as extended states^{20,21}, although they do not form an energy band.

A recent attempt to introduce long-range correlations in 1D systems utilizes $\{\varepsilon_i\}$ drawn from a continuous probability distribution²². Because in this case the fluctuations in the diagonal energies (and thus the disorder in the system) increase with the system size, this work was limited to finite systems, and found

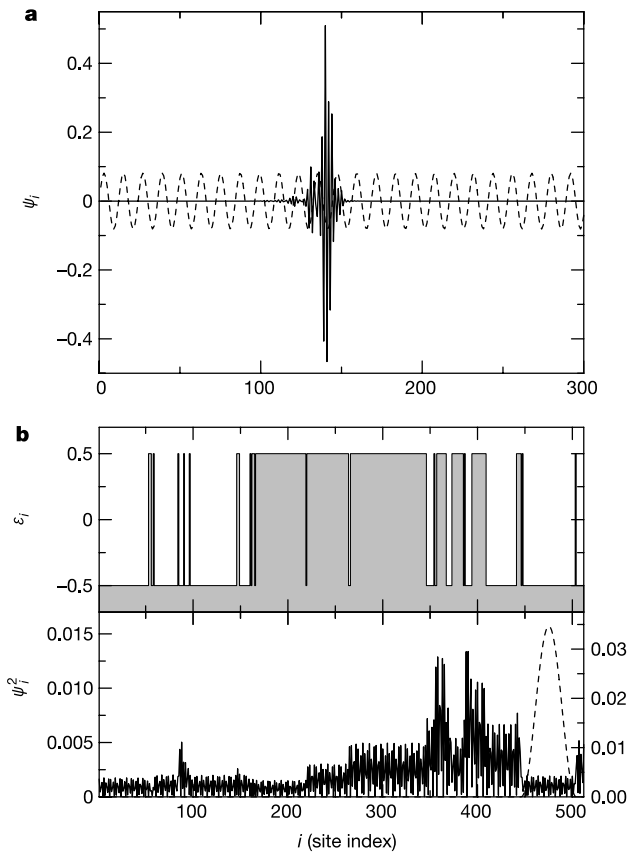


Figure 1 Wavefunctions for periodic, disordered and correlated-disordered chains. **a**, Typical wavefunctions for a periodic binary chain (dashed line) spreading over the whole system and for a random binary alloy (solid line) localized in a small region. Both are calculated for a system of size $N = 300$ atoms, and for $\varepsilon_A = 1/2$ and $\varepsilon_B = -1/2$. The wavefunctions are obtained by means of numerical diagonalization of the hamiltonian (equation (1)). **b**, Top panel: energy profile realization for a binary chain of $N = 512$ atoms with correlation exponent $\alpha = 1.2 < \alpha_c$. Bottom panel: two typical wavefunctions corresponding to a localized and a nearly extended state obtained by numerical diagonalization of the hamiltonian for the profile shown in the top panel. The localized state (dashed line) which corresponds to an energy $E = -1.5372$, randomly chosen outside the conducting band $[-1.5, 1.5]$, is different from zero only inside a cluster of identical B atoms with $\varepsilon_B = -1/2$ shown in the top panel. Out of this cluster the wavefunction becomes zero, indicating that the electron is localized and confined inside the cluster — a behaviour typical of disordered systems. However, for the nearly extended state, which correspond to an energy $E = 0.05421$ inside the central band, a very strong coupling mechanism among different correlated clusters is observed, and the wavefunction (solid line) is clearly different from zero almost everywhere, indicating delocalization of the electron even when $\alpha < \alpha_c$. This coupling mechanism is responsible for the electronic delocalization, and it only occurs in the presence of correlations and for electron states corresponding to the central energy band. The vertical scale on the left axis corresponds to the extended wavefunction, and the right axis to the localized one.

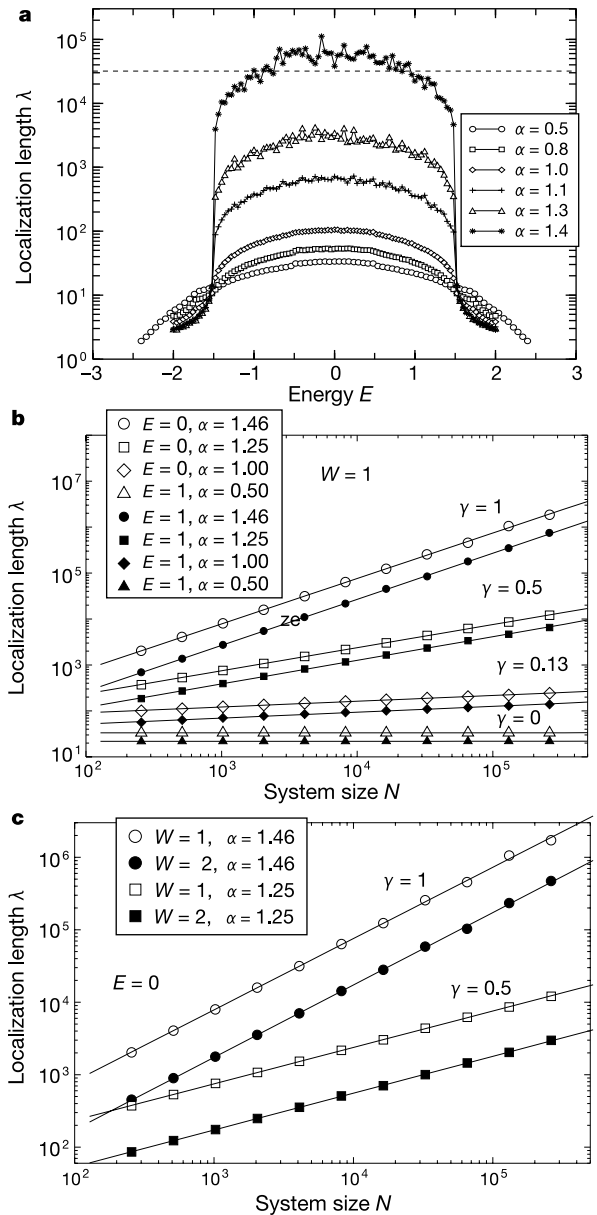


Figure 2 Localization length behaviour in correlated-disordered chains. **a**, Behaviour of the localization length λ as a function of energy E for several values of the correlations. The horizontal dashed line corresponds to the fixed system size $N = 2^{15} \approx 3 \times 10^4$. For $\alpha \approx 1.4$ we find a threshold above which extended (delocalized) states with $\lambda \approx N$ exist in the central region of the energy band $[\varepsilon_A - 2t, \varepsilon_A + 2t] \cap [\varepsilon_B - 2t, \varepsilon_B + 2t]$. As $t = 1$, $\varepsilon_A = -\varepsilon_B = 1/2$, this energy band is $[-1.5, 1.5]$. **b**, Scaling relation $\lambda \propto N^\gamma$ for different values of the correlation exponent α . This behaviour is identical for all energy values in the central band $-1.5 < E < 1.5$. We show results for $E = 0$ and $E = 1$ while keeping the disorder parameter fixed at $W = 1$. To avoid numerical fluctuations, λ is averaged in a small window of width $\Delta E = 0.1$ around each energy value. N ranges from 2^8 to 2^{18} atoms. The solid lines correspond to fits of the type $\lambda \propto N^\gamma$, from where the exponent γ is obtained. For a random binary chain with $\alpha = 0.5$, λ remains constant for all values of N and correspondingly $\gamma = 0$. For a correlated binary chain with $\alpha > \alpha_c$ we find $\gamma = 1$, indicating that above the ‘critical point’ of the metal–insulator transition the localization length is of the size of the system N . Note that, although λ is smaller for $E = 1$ compared to $E = 0$ for all α , the exponent γ remains unchanged. Thus the observed quantum insulator–metal transition is indeed valid in the thermodynamic limit and for a broad energy band. **c**, λ as a function of N for different degrees of disorder $W = |\varepsilon_A - \varepsilon_B|$ and for fixed $E = 0$. Although the system with higher disorder (filled symbols) exhibits shorter localization lengths for all system sizes, the scaling property $\lambda \propto N^\gamma$ is not altered with W , for all values of α . Thus independently of W (provided there is an intersection between the bands of A and B) the system exhibits an insulator–conductor transition for $\alpha > \alpha_c$ in the thermodynamic limit.

extended states with increasing strength of the correlations only for $E \rightarrow 0$. In order to control the disorder and to maintain the spread of the density of states, a normalization to unit standard deviation is needed, which ultimately eliminates the correlations in the ε_i and produces a series of energies that tend to be constant²³.

We propose a binary 1D model of correlated disorder, which generates a broad energy band of extended states and leads to conductivity even in the thermodynamic limit of large system size. For a perfectly ordered 1D system with atoms of only type A, the energy spectra of the solid form a single energy band $[\varepsilon_A - 2t, \varepsilon_A + 2t]$, $W/t = 0$, and we have a perfect conductor. For a random binary solid, $W \equiv |\varepsilon_A - \varepsilon_B|$. The closer the values of ε_A and ε_B are, the smaller is the ratio W/t , and the conducting properties of the disordered solid become similar to those of the perfectly ordered solid. The larger W is, the stronger is the localization. We choose $W/t = 1$, and we have strong localization in the case of pure disorder. For long-range power-law correlations in the diagonal energies ε_i , we still have $W/t = 1$. Thus, we are able to obtain extended states by keeping the level of disorder fixed, which allows conducting behaviour in the thermodynamic limit and in a broad energy band.

We generate binary sequences with long-range correlations using the modified Fourier filter method²⁴, and we use these sequences as the $\{\varepsilon_i\}$ in the binary chain. The algorithm generates random noise in the frequency domain, multiplies this noise by a power-law with the desired exponent, and then Fourier-transforms the signal back into real space. As we must generate a binary sequence, we consider a transformation which maps any positive value of the correlated series into ε_A , and any negative value into ε_B . Such a mapping can change the correlation properties of the series, and therefore the correlations are not properly quantified by the power-law exponent in the original correlated series. To quantify the correlations in the final binary sequence, we calculate the scaling exponent α using detrended fluctuation analysis^{25,26}. A value of $\alpha = 0.5$ corresponds to an uncorrelated random sequence (white noise). If $\alpha < 0.5$, the binary sequence is anticorrelated, while if $\alpha > 0.5$, there are positive long-range correlations (Supplementary Information).

Next, we study the effect of long-range power-law correlations on the localization properties of disordered binary chains. We calculate λ using the transfer matrix method¹¹. The Schrödinger equation for the hamiltonian (equation (1)) becomes

$$E\psi_n = t\psi_{n-1} + \varepsilon_n\psi_n + t\psi_{n+1} \quad (2)$$

where E is the energy corresponding to the electron wavefunction, $|\psi_n|^2$ is the probability of finding an electron at site n in the chain, and $\varepsilon_n = \varepsilon_A$ or ε_B (Fig. 1b). Using the transfer matrix method, we write equation (2) in the recursive form:

$$M_n \begin{pmatrix} \psi_n \\ \psi_{n-1} \end{pmatrix} = \begin{pmatrix} \psi_{n+1} \\ \psi_n \end{pmatrix}, \quad M_n = \begin{pmatrix} E - \varepsilon_n & -t \\ t & 0 \end{pmatrix} \quad (3)$$

The localization length $\lambda(E)$ is then defined by

$$\frac{1}{\lambda(E)} = \lim_{N \rightarrow \infty} \frac{1}{N} \ln \left| \frac{\psi_N}{\psi_0} \right| \quad (4)$$

where N is the chain length. We choose $\psi_0 = \psi_1 = 1/\sqrt{2}$. For every realization of the potential we apply equation (3) to obtain ψ_N , and we calculate $\lambda(E)$ from equation (4). We average λ over a large ensemble of realizations for a fixed system size N and fixed value of the energy E . We repeat this procedure for different values of N and E .

Our results for λ show that correlations in the sequence of $\{\varepsilon_i\}$ in the binary chain strongly modify the localization properties of the wavefunctions (Fig. 2). For $\alpha = 0.5$ and fixed N , we obtain the shortest localization length as a function of E , which corresponds to the result for a random binary alloy¹⁹. We find that long-range correlations ($\alpha' > 0.5$) change drastically the localization proper-

ties of the electronic states. When we increase the long-range correlations in the chain, λ increases in the central region of the energy band, and eventually, for a fixed N , we find a critical value α_c above which $\lambda > N$, which corresponds to extended wavefunctions (Fig. 2a). The energy region $[\varepsilon_A - 2t, \varepsilon_A + 2t] \cap [\varepsilon_B - 2t, \varepsilon_B + 2t]$ in which this increase of λ takes place is precisely the overlap of the energy bands corresponding to ordered chains formed only by A atoms and only by B atoms, so we obtain a broad band of extended states. In the random dimer model^{20,21}, there is only a single energy value that corresponds to an extended state, implying that the probability of an electron having precisely this energy is small for

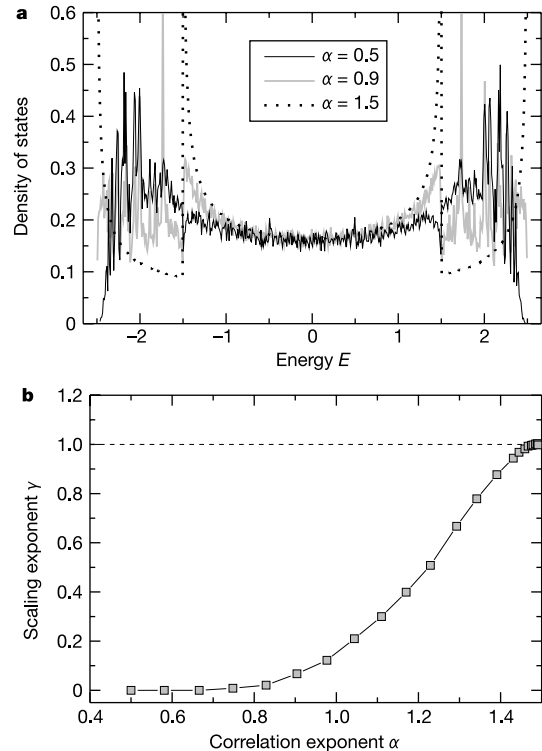


Figure 3 Density of states and scaling exponent as a function of the correlations. **a**, Density of states as a function of energy for three different values of the correlation exponent α and for $N = 2^{16}$. As the strength of the correlations increases (that is, larger values of α), there is (1) a migration of electron states to the central region $-1.5 < E < 1.5$ where the delocalization takes place, and (2) the density of states approaches the form expected for a system of two semi-infinite clusters of type A and B—four peaks with central conducting region (Fig. 2a). This reflects the fact that with increasing α the probability of finding large clusters of type A or type B in the system with correlated disorder also increases, although clusters of all sizes are present (Supplementary Information). We find that the delocalized states (conducting electrons) form a substantial fraction of the total number of electron states, and that this fraction increases with α —52.5% for $\alpha = 0.5$, 57.8% for $\alpha = 0.9$ and 66.6% for $\alpha = 1.5$. For given α the fraction of delocalized states remains the same with varying system size N . Thus for $\alpha > \alpha_c$ the majority of electron states are extended and the system exhibits metallic behaviour. **b**, The scaling exponent γ as a function of α . The value $\gamma = 1$, which corresponds to extended (delocalized) electron states in the limit $N \rightarrow \infty$, is obtained for $\alpha \approx 1.45$, suggesting a critical point at which there is transition from insulating to metallic behaviour. Note that the observed enhanced conductivity and quantum metal–insulator transition in 1D binary systems with long-range correlations is not a trivial consequence of the presence of long repetitive sequences of atoms of type A followed by long repetitive sequences of atoms B, provided the energy bands of A and B overlap. Even in the presence of very strong correlations, there is a mixture of both small and large clusters of type A and B. This is not surprising, as the correlations we introduce in the system are power-law long-range correlations, and thus the energy profile is self-similar on different size scales. Our results suggest that this self-similar organization of clusters of different size can lead to extended states and metal–insulator transition.

finite systems, and becomes zero in the thermodynamic limit of large system size. In contrast, long-range power-law correlations produce a broad conducting energy band. Moreover, the population of this central band increases with the correlations, ranging from

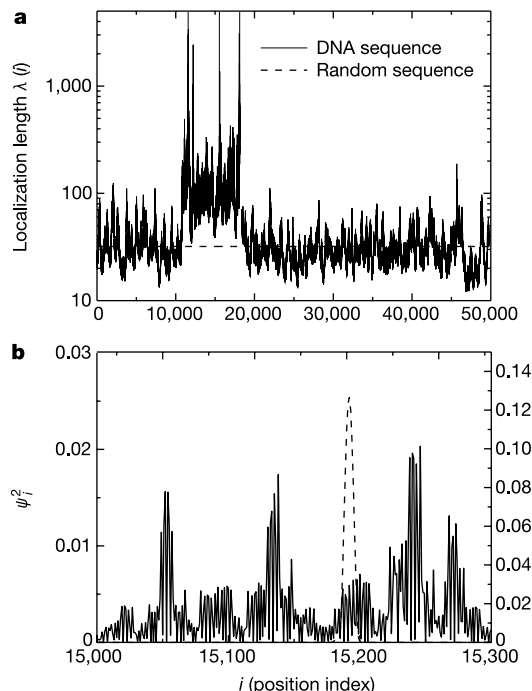


Figure 4 Results in DNA. **a**, λ as a function of the position in a DNA sequence. We present the following experiment: (1) We select a DNA sequence corresponding to the first 50,000 nucleotides of the largest contig (a perfectly sequenced region without gaps) of human chromosome 22 (NT_011520) retrieved from the National Center for Biotechnology Information (NCBI). (2) We map this sequence onto a binary alphabet so that nucleotides A and T gives $\varepsilon_A = 1/2$ and nucleotides C and G give $\varepsilon_B = -1/2$ to construct the diagonal energies in the hamiltonian (equation (1)). (3) We divide the sequence into overlapping subsequences of size $N = 300$, one for each nucleotide in the original DNA sequence (that is, for nucleotide i we consider the subsequence between $[i, i + 299]$), and we calculate λ for $E = 0$ and for each i . (4) We plot λ as a function of the sequence position i . Note that there is a certain region of non-repetitive DNA comprising $N \approx 8,000$ nucleotides and characterized by strong long-range correlations with $\alpha \approx 0.9$. For this region, we find that λ is on average ten times greater compared to the neighbouring segments of DNA where there are no strong correlations, and compared to a surrogate random sequence (horizontal dashed line). Thus electron transport in DNA can be clearly enhanced owing to long-range correlations, leading to conductivity over DNA segments comprising hundreds or even thousands of nucleotides. Looking for segments of DNA from the human chromosome 22 which have similar correlation properties and enhanced electron transport, we found that the biological function of these segments is not provided and is not known. Perhaps the patterns formed by segments with particular conduction (correlation) properties can provide a clue to understanding their biological function. **b**, Two wavefunctions corresponding to a subsequence of length $N = 300$ located at $i = 15,000$, as described in **a**. The wavefunctions are obtained by numerical diagonalization of the hamiltonian in equation (1) with diagonal energies corresponding precisely to these 300 nucleotides. The solid line (left vertical axis) represents a nearly extended state corresponding to an energy value $E = -0.17753$ from the central band $[-1.5, 1.5]$, and the dotted line (right vertical axis) shows a strongly localized state corresponding to $E = 1.55761$ outside the central band. Whereas the localized state is confined within ~ 30 base pairs, the extended state allows for an electron transport throughout the sequence. Moreover, because λ exceeds the size of the segments with correlated disorder, the electron wavefunction overlaps with parts of the neighbouring uncorrelated disordered segments and can affect their conducting properties, thus facilitating important genomic functions²⁷. Note the similarity between this plot obtained for DNA and the wavefunction in Fig. 1b obtained for a computer-generated correlated binary sequence.

52% for $\alpha = 0.5$ to 67% for $\alpha = 1.5$ (Fig. 3a). Thus, the probability of an electron propagating throughout the system, which is proportional to the width of the conducting energy band, is not small, and the conducting behaviour resembles a metal.

To speak properly of extended states, we need the condition $\lim_{N \rightarrow \infty} (\lambda/N) = \text{constant} \neq 0$. If this condition is not fulfilled, λ becomes negligible in comparison to N in the thermodynamic limit $N \rightarrow \infty$, and the wavefunction is therefore localized. Hence, we next study the behaviour of λ as a function of N for different values of the correlations. As λ is also a function of the electron energy E , we choose $E = 0$ —that is, the central energy of the band. To avoid excessive numerical fluctuations, we average λ in a small energy window around $E = 0$, specifically the interval $[-0.05, 0.05]$. We find a power-law relation between λ and N , $\lambda \propto N^\gamma$ (Fig. 2b, c), where the exponent γ depends on the correlation exponent α (Fig. 3b). Our results for the values of $\lambda(E)$ (Fig. 2a) indicate a nearly flat plateau covering the central energy band, suggesting that a similar relation between λ and N can be expected for every other value of E in this energy interval. Indeed, for all $E \in [-1.5, 1.5]$ we find the same values of γ (Fig. 2b).

Further, we find that γ is also independent of W (Fig. 2c). As W is fixed, introducing power-law correlations in the chain leads to re-ordering of the diagonal energy values ε_i to produce clusters the sizes of which are power-law distributed (Supplementary Information), and at the same time not changing the level of disorder when N varies. We find that a change in W modifies the values of λ : λ decreases with increasing W . But owing to the power-law correlations, the fraction λ/N remains constant when $N \rightarrow \infty$ (note the constant vertical shift in Fig. 2c), while the scaling relation, $\lambda \propto N^\gamma$, does not change with W . Thus, in the thermodynamic limit, this scaling behaviour is independent of the level of disorder.

For values of $\alpha \approx 0.5$ we find that $\gamma \approx 0$, so we have localized states. As α increases, γ also increases; for $\alpha \geq 1.45$ we obtain $\gamma = 1$ within our error bars (Fig. 3b). Therefore, at $T = 0$ and in the limit $N \rightarrow \infty$, for $\alpha < 1.45$, the system behaves as an insulator, while for $\alpha > 1.45$ the system behaves as a conductor within a broad energy band. For $\alpha = \alpha_c = 1.45$, we find a new ‘critical point’ at which a quantum metal–insulator transition occurs.

These findings can be used to better understand the conduction properties in 1D disordered binary solids, which are important for practical applications such as the construction of nanoscopic electronic devices⁹. Electrical conduction in biological macromolecules, in particular DNA, has also attracted much attention. Recent work has shown that electrons or holes are responsible for the electrical current in DNA^{7–10}. Moreover, it has been observed that certain mutation repairs occur in natural DNA by means of electrical current transport along the molecule²⁷. Experiments of DNA conductivity find conducting^{7,28}, semiconducting⁹ and insulating²⁹ behaviour, due perhaps to the different type of DNA sequences considered—random, repetitive or correlated. An open question is the influence of the ‘ordering’ (that is, correlations) of the DNA nucleotides on the conduction properties of the sequence.

In DNA, correlations with exponent $\alpha \approx \alpha_c$ do not exist. Typical values of α in DNA range between 0.6 and 0.9 (refs 5, 6), which are below α_c but are significantly greater than the value of 0.5 for purely random sequences. Our results show that such long-range correlations can strongly affect the electronic transport over relatively large distances. In particular, we consider a sequence of 50,000 nucleotides of the largest contig of human chromosome 22 (Fig. 4). We find that for non-repetitive regions with long-range correlations, the localization length obtained with our model when the DNA is mapped onto a binary chain is between one and two orders of magnitude greater (Fig. 4a) than that expected for a random sequence (Fig. 2a, b). This behaviour is also reflected in the fact that we obtain nearly extended wavefunctions for all energy values from the central band of the DNA sequence (Fig. 4b). Recent experiments show that random DNA is not a good conductor²⁹, whereas

repetitive DNA conducts quite well³⁰—both cases are in agreement with our model (Figs 2 and 3). Further, our results suggest that in non-repetitive long-range correlated regions of DNA, electrons can propagate over average distances of ~ 300 nucleotides, and that a fraction can propagate over distances of more than 1,000 nucleotides. In fact, the DNA segment in Fig. 4a, where our model predicts very good conducting behaviour, is even longer—extending to $\sim 8,000$ nucleotides. This conducting behaviour does not imply that correlated DNA is a macroscopic conductor, but rather that electronic transport at moderate distances can be found at $T = 0$. This distance range ($\sim 1 \mu\text{m}$) is the focus of the above-mentioned experiments.

In summary, we find that long-range correlations change the localization properties of 1D disordered binary solids. We show that the localization length of the electron wavefunction is greatly increased by long-range correlations. In addition, for correlations stronger than a certain threshold, we find in the thermodynamic limit a broad energy band of extended states, and therefore a conducting phase. Thus, although still disordered, the 1D system can behave as a conductor; in contrast to the traditional theory which is applicable only for uncorrelated disorder. The threshold in the control parameter (the value of the correlation exponent) corresponds to a ‘critical point’ at which a metal–insulator transition takes place. These findings may be of importance for elucidating the electronic transport and the biological function of DNA segments with different types of correlations, as well as for the design of nanoscopic devices. □

Received 1 May; accepted 28 June 2002; doi:10.1038/nature00948.

1. Ashcroft, N. W. & Mermin, N. D. *Solid State Physics* (Holt-Saunders, London, 1976).
2. Anderson, P. W. Absence of diffusion in certain random lattices. *Phys. Rev.* **109**, 1492–1505 (1958).
3. Kramer, B. & MacKinnon, A. Localization: theory and experiment. *Rep. Prog. Phys.* **56**, 1469–1564 (1993).
4. Janssen, M. Statistics and scaling in disordered mesoscopic electron-systems. *Phys. Rep.* **295**, 1–91 (1998).
5. Peng, C.-K. *et al.* Long-range correlations in nucleotide sequences. *Nature* **356**, 168–170 (1992).
6. Holste, D., Grosse, I. & Herzel, H. Statistical analysis of the DNA sequence of human chromosome 22. *Phys. Rev. E* **64**, 041917 (2001).
7. Fink, H. V. & Schönenberger, C. Electrical conduction through DNA molecules. *Nature* **398**, 407–410 (1999).
8. Henderson, P. T., Jones, D., Hampikian, G., Kan, Y. & Schuster, G. B. Long distance charge transport in duplex DNA: The phonon-assisted polaron-like hopping mechanism. *Proc. Natl Acad. Sci. USA* **96**, 8353–8358 (1999).
9. Porath, D., Bezryadin, A., de Vries, S. & Dekker, C. Direct measurement of electrical transport through DNA molecules. *Nature* **403**, 635–638 (2000).
10. Hjort, M. & Stafstrom, S. Band resonant tunnelling in DNA molecules. *Phys. Rev. Lett.* **87**, 228101 (2001).
11. MacKinnon, A. & Kramer, B. One-parameter scaling of localization length and conductance in disordered systems. *Phys. Rev. Lett.* **47**, 1546–1549 (1981).
12. Hofstadter, E. & Schreiber, M. Statistical properties of the eigenvalue spectrum of the three-dimensional Anderson hamiltonian. *Phys. Rev. B* **48**, 16979–16985 (1993).
13. Kravchenko, S. V., Simonian, D., Sarachik, M. P., Mason, W. & Furneaux, J. E. Electric-field scaling at a $B = 0$ metal–insulator transition in 2 dimensions. *Phys. Rev. Lett.* **77**, 4398–4941 (1996).
14. Weymer, A. & Janssen, M. Localization length exponent, critical conductance distribution and multifractality in hierarchical networks models for the quantum Hall-effect. *Ann. Phys. (Leipzig)* **7**, 159–173 (1998).
15. Schweitzer, L. & Zharekeshev, I. Kh. Scaling of level statistics and critical exponent of disordered 2-dimensional symplectic systems. *J. Phys. Condens. Matter* **9**, L441–L445 (1997).
16. Mott, N. F. & Twose, W. The theory of impurity conduction. *Adv. Phys.* **10**, 107–163 (1961).
17. Abou-Chacra, R., Anderson, P. W. & Thouless, D. J. A selfconsistent theory of localization. *J. Phys. C*, **6**, 1734–1752 (1973).
18. Ishii, K. & Matsuda, H. Localization of normal modes and energy transport in the disordered harmonic chain. *Prog. Theor. Phys. Suppl.* **45**, 56–86 (1970).
19. Davids, P. S. Lyapunov exponents and transfer-matrix spectrum of the random binary alloy. *Phys. Rev. B* **52**, 4146–4155 (1995).
20. Dunlap, D. H., Wu, H.-L. & Phillips, P. Absence of localization in a random-dimer model. *Phys. Rev. Lett.* **65**, 88–91 (1990).
21. Phillips, P. & Wu, H.-L. Localization and its absence: a new metallic state for conducting polymers. *Science* **252**, 1805–1812 (1991).
22. de Moura, F. A. B. F. & Lyra, M. L. Delocalization in the 1D Anderson model with long-range correlated disorder. *Phys. Rev. Lett.* **81**, 3735–3738 (1998).
23. Kantelhardt, J. W., Russ, S., Bunde, A., Havlin, S. & Webman, I. Comment on delocalization in the 1D Anderson model with long-range correlated disorder. *Phys. Rev. Lett.* **84**, 198–201 (2000).
24. Makse, H. A., Havlin, S., Schwartz, M. & Stanley, H. E. Method for generating long-range correlations for large systems. *Phys. Rev. E* **53**, 5445–5449 (1996).
25. Hu, K., Ivanov, P. Ch., Chen, Z., Carpena, P. & Stanley, H. E. Effect of trends on detrended fluctuation analysis. *Phys. Rev. E* **64**, 011114 (2001).

26. Chen, Z., Ivanov, P. Ch., Hu, K. & Stanley, H. E. Effect of nonstationarities on detrended fluctuation analysis. *Phys. Rev. E* **65**, 041107 (2002).
27. Dandliker, P. J., Holmlin, R. E. & Barton, J. K. Oxidative thymine dimer repair in the DNA helix. *Science* **275**, 1465–1468 (1997).
28. Kasumov, A. I. *et al.* Proximity-induced superconductivity in DNA. *Science* **291**, 280–282 (2001).
29. de Pablo, P. J. *et al.* Absence of dc-conductivity in λ -DNA. *Phys. Rev. Lett.* **85**, 4992–4995 (2000).
30. Yoo, K.-H. *et al.* Electrical conduction through poly(dA)–poly(dT) and poly(dG)–poly(dC) DNA molecules. *Phys. Rev. Lett.* **87**, 198102 (2001).

Supplementary information accompanies the paper on Nature’s website (<http://www.nature.com/nature>).

Acknowledgements

We thank L. Cruz for discussions, and the Spanish Ministerio de Educación y Cultura and NIH/National Center for Research Resources for support.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.C. (e-mail: pcarpa@ctima.uma.es or pcarpa@uma.es).

Near-infrared sensitivity enhancement of photorefractive polymer composites by pre-illumination

Erwin Mecher*, Francisco Gallego-Gómez†, Hartwig Tillmann‡, Hans-Heinrich Hörhold‡, Jan C. Hummelen§ & Klaus Meerholz*†

* Chemistry Department and Center for Nanoscience (CeNS), University of Munich, Butenandstrasse 11, 81377 Munich, Germany

‡ Organic and Macromolecular Chemistry, University of Jena, Humboldtstrasse 10, 07743 Jena, Germany

§ Stratingh Institute and Materials Science Center, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Among the various applications for reversible holographic storage media^{1,2}, a particularly interesting one is time-gated holographic imaging (TGHI)^{3–5}. This technique could provide a noninvasive medical diagnosis tool, related to optical coherence tomography^{6,7}. In this technique, biological samples are illuminated within their transparency window with near-infrared light, and information about subsurface features is obtained by a detection method that distinguishes between reflected photons originating from a certain depth and those scattered from various depths. Such an application requires reversible holographic storage media with very high sensitivity in the near-infrared. Photorefractive materials, in particular certain amorphous organic systems, are in principle promising candidate media, but their sensitivity has so far been too low, mainly owing to their long response times in the near-infrared. Here we introduce an organic photorefractive material—a composite based on the poly(arylene vinylene) copolymer TPD-PPV⁸—that exhibits favourable near-infrared characteristics. We show that pre-illumination of this material at a shorter wavelength before holographic recording improves the response time by a factor of 40. This process was found to be reversible. We demonstrate multiple holographic recording with this technique at video rate under practical conditions.

† Present address: Physical Chemistry Department, University of Cologne, Luxemburgerstrasse 116, 50939 Cologne, Germany.

Optical coherence tomography (OCT) has some shortcomings concerning the recording time, which could be overcome by using TGHI. The fundamental physical processes are identical for both techniques; the material is illuminated with low-coherence continuous-wave light or ultra-short laser pulses. Photons are mostly scattered, but a small percentage are reflected on material-characteristic features (so-called ‘ballistic’ photons). The scattered and reflected light is collected and fed into an interferometer (Michelson interferometer in OCT, holographic imaging system in TGHI). Because only the ballistic photons which come from a certain depth are coherent with the reference beam, an optical delay line can be used to determine the depth of the features to be recorded. The image contrast results from a combination of absorption and scattering features.

Photorefractive (PR) materials are well suited for TGHI, as the PR effect allows for recording holograms with high index modulation amplitudes even at low light levels owing to its time-integrating character. Also, the PR effect is reversible, that is, previously recorded holograms are erased or overwritten, a feature that would be mandatory for a commercial TGHI system. Photorefractivity requires photoconductivity and electro-optical response. In PR polymer composites these functions are achieved simply by mixing the appropriate functional components^{9,10}. Most commonly, a sensitizing and nonlinear-optical chromophores are doped into a hole-conducting polymer host. To write a hologram the material is pre-poled by an external poling field E_0 and then illuminated by two coherent intersecting laser beams. In the bright regions of the resulting interference pattern absorption and field-

enhanced exciton dissociation take place. The mobile charge carriers (mostly holes) are redistributed and become trapped in the dark regions, leading to an internal space-charge field E_{sc} (ref. 11). The superposition of E_0 and E_{sc} acts on the NLO chromophores, modulating the refractive index of the bulk to replicate the interference pattern^{12,13}.

A high sensitivity S is required to record a hologram with a sufficiently large external diffraction efficiency $\eta_{\text{ext}}(t_{\text{exp}})$ at low light levels (total external write-beam intensity $I_{\text{WB,ext}}$) and after a short exposure time t_{exp} . A technically widely used definition is¹⁴ (see Methods):

$$S = \frac{\sqrt{\eta_{\text{ext}}(t_{\text{exp}})}}{I_{\text{WB,ext}} t_{\text{exp}}} \quad (1)$$

In organic PR materials a high sensitivity can be achieved by: (1) generating the highest possible PR space-charge field E_{sc} ; (2) optimizing the EO chromophores¹⁵; (3) decreasing the absorption losses; and/or finally (4) decreasing the exposure time by increasing the recording speed. Here we report on the new photo-physical phenomenon of ‘gating’ in organic PR materials, which simultaneously tackles problems (1), (3) and (4) to yield unmatched near-infrared sensitivity in TPD-PPV-based PR devices.

$\eta_{\text{ext}}(t)$ was measured by four-wave mixing in a typical tilted geometry⁹ with a wavelength of 830 nm (see Methods). The recording dynamics of the TPD-PPV-based material (see Methods for composition) was found to depend sublinearly on the recording intensity (not shown). This indicated that the grating build-up was limited by the generation of charge carriers.

Our method of speeding up the near-infrared recording process was specifically to pre-illuminate the material to provide carriers before the writing starts. Under these conditions the redistribution of the carriers (transport and trapping) might become the limiting step. Some attempts in this direction have already been reported in the literature: Silence *et al.*¹⁶ found that pre-illumination irreversibly led to a slower response, but improved refractive index modulation amplitude Δn . The effect was referred to as ‘optical trap activation’. Similarly, Herlocker *et al.* also reported a slowing of the response time owing to accumulation of traps¹⁷. Finally, Wolff *et al.* found no effect of pre-illumination on the recording process in their material¹⁸. Instead, when illuminated during recording the recording became faster, an effect referred to as ‘optical trap filling’. Unfortunately, a strong loss in diffraction efficiency was anticipated. In our material, pre-illumination led to an acceleration of the recording process, unlike in all previous attempts. The pre-illumination effects were more pronounced when an independent light pulse of shorter wavelength was applied. In the context of this paper we refer to this procedure as ‘gating’, however, we point out that the mechanism discussed here is different from ‘gating’ in inorganic PR materials^{19,20}.

Figure 1a shows holographic recording traces for gating with $\lambda = 633$ nm light, using various intensities I_g and constant pulse duration t_g . A strong decrease of the response time by more than one order of magnitude and also a slight increase in η is observed.

Our explanation is the following. Gating leads to a spatially uniform distribution of charge carriers, which are absent in the pristine material (Fig. 2a). The carrier generation is much more efficient at shorter than at longer wavelengths, so gating is more efficient when short-wavelength light is used. The increase in η indicates that the number density of ‘PR traps’ N_t is increased by the pre-produced charges^{16,21} and therefore E_{sc} can reach higher values. This interpretation is supported by measurements of the PR gain coefficient Γ , which was found to be larger in the non-illuminated (10 cm^{-1}) than in the gated samples (5 cm^{-1}) despite the increased Δn in the latter case, consistent with a higher trap density ($N_t \approx 2 \times 10^{17} \text{ cm}^{-3}$ for gated compared with 10^{17} cm^{-3} for non-gated; see Methods). Beam coupling can lead to image distortions, so the

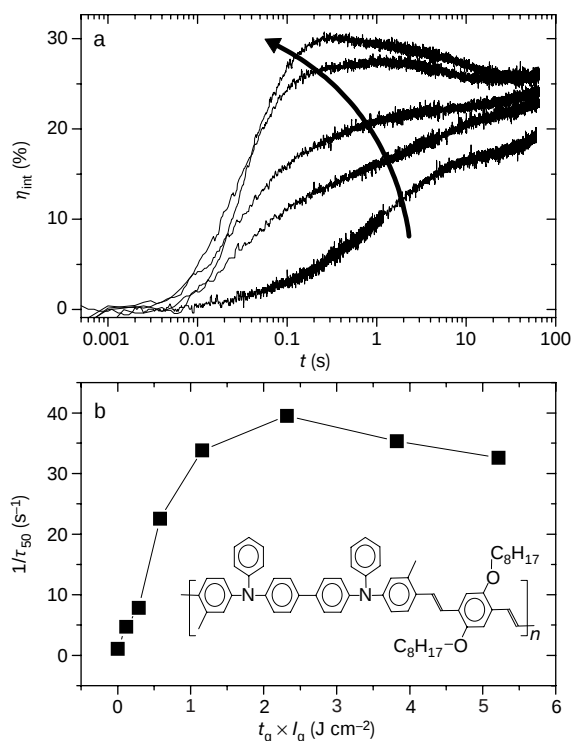


Figure 1 Holographic recording dynamics of the TPD-PPV-based composite for different gate intensities and chemical structure of TPD-PPV. **a**, Temporal evolution of the internal diffraction efficiency without gating (lowest curve) and after gate pulses with an intensity $I_g = 0.29, 0.58, 2.3$ and 5.2 W cm^{-2} , respectively. Total write-beam intensity $I_{\text{WB,ext}} = 3.27 \text{ W cm}^{-2}$; gating time $t_g = 955$ ms; and external field $E_{\text{ext}} = 60 \text{ V } \mu\text{m}^{-1}$. The arrow indicates increasing gate intensity. **b**, Dependence of the inverse response times τ_{50}^{-1} on the gating fluence. τ_{50} is the time necessary to reach 50% of the quasi-steady-state value of the diffraction efficiency (at 60 s). Inset, chemical structure of TPD-PPV.

low gain coefficient found for this material is favourable for diffraction-based applications such as TGHI.

For high gating exposures intermediate diffraction maxima are observed in the temporal dependence (see Fig. 1a, top curve). Because temporal diffraction minima upon erasure are not observed, effects from competing hole and electron gratings as reported in ref. 22 can be excluded as an explanation. Instead, the occurrence of a temporal diffraction maximum indicates that E_{SC} is higher at intermediate times than in the quasi-steady state. This is the result of the advancing spatial charge separation (redistribution) on the one hand (higher E_{SC}) and a decreasing PR trap concentration caused by recombination on the other (lowering E_{SC}) as illustrated in Fig. 2b (time t_2).

The most prominent effect of the gating process is a strong increase of the recording speed. As a metric of comparison for the recording speed we plot $1/\tau_{50}$, the inverse of the time necessary to reach 50% of the quasi-steady-state value of the diffraction efficiency (Fig. 1b). Overall, the speed-up between gated and non-gated recording amounts to a factor of 40. This is an indication that

indeed the carrier generation no longer limits the speed of the entire recording process. Instead the redistribution of carriers becomes the limiting process. This interpretation is supported by the finding that the gated recording speed at 830 nm is only about twice as slow as at a write wavelength of 633 nm, whereas the difference is much larger (by a factor of 80) for non-gated samples. In this context, we cannot distinguish whether or not the carrier mobility was affected by the gating.

The carrier distribution produced by the gating process is expected to relax over time due to recombination. To verify this a temporal delay t_d was introduced between applying the gate beam and starting the writing process. We found that the recording slows down exponentially, reaching a relaxed state (equal to non-gated samples) after $t_d \approx 50$ s. From this we conclude that the recording speed correlates with density of pre-existing charge carriers.

With increasing gate fluence, the number density of free charge carriers will pass through a maximum that is due to trapping, which explains the reduction in recording speed for high pre-exposure doses (Fig. 1b). Because we observe faster recording speed by pre-illumination as well as by illumination during recording, unlike Wolff *et al.*¹⁸, this indicates that the carriers remain 'free' for a rather long time. These considerations are backed by transient measurements of the photocurrent using experimental conditions similar to those used here (L. Kulikovskiy, E.M., K.M. and D. Neher, manuscript in preparation) and are further supported by redox-chemical

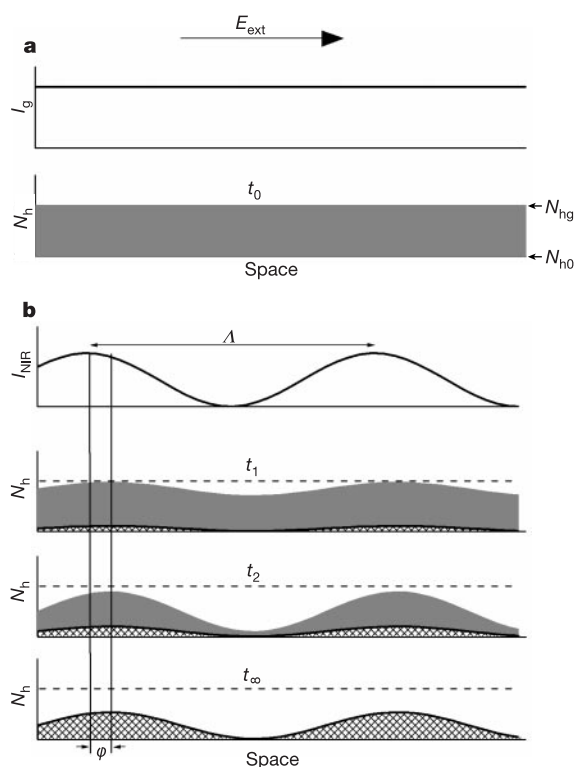


Figure 2 Illustration of the gating mechanism. **a**, A pristine sample exhibits a vanishing, spatially uniform charge carrier distribution N_0 . Uniform pre-illumination with a gate beam of intensity I_g (top) increases the carrier density to N_g (bottom). For reasons of electroneutrality, the number densities of holes and electrons are equal. In TPD-PPV holes are much more mobile than electrons, so we show the distribution N_h of holes only. **b**, During the writing procedure the interference pattern with the intensity I_{NIR} (top, grating period Λ) spatially modulates the hole density N_h . Snapshots of the situation before steady state is reached are shown for two write times $t_1 < t_2$ (upper/lower middle) and the quasi-steady-state value t_∞ (bottom). Without gating free holes are generated in the bright regions and redistributed (black cross-hatched area). The modulation amplitude increases monotonously in time and finally reaches the quasi-steady-state value at t_∞ . The recording process is relatively slow. By contrast, with gating the modulation (grey filled area) is 'carved' in the initial hole density (dashed line). At intermediate times (t_2) the modulation can assume larger values than in the quasi-steady-state value (t_∞), which is identical to the case without gating. The gated recording process is relatively fast. For simplicity, this illustration assumes a constant spatial relocation φ of the carriers relative to the incident interference pattern.

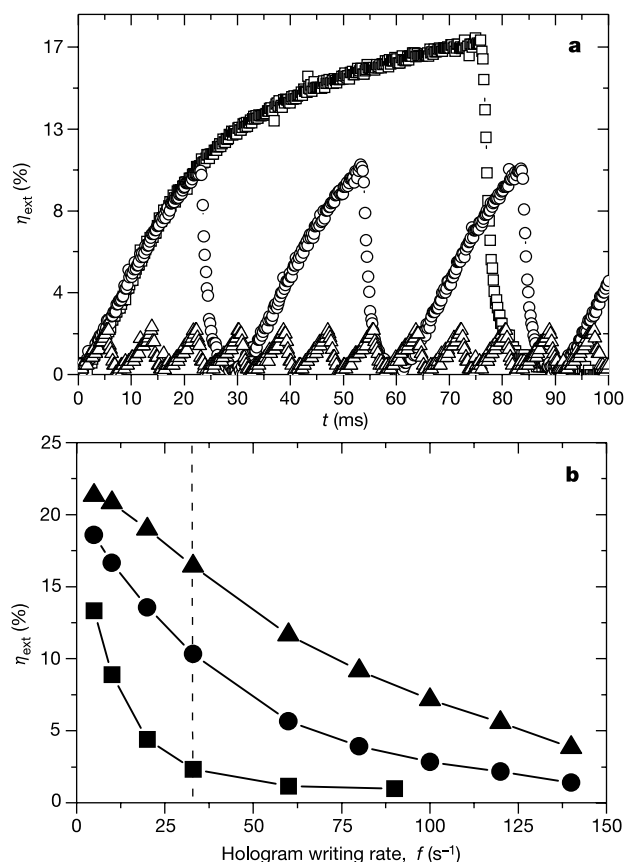


Figure 3 Gated holographic recording dynamics of the TPD-PPV-based composite for different write rates and write intensities. **a**, Time dependence of the external diffraction efficiency for a hologram write rate of 10 Hz (squares), 33 Hz (circles) and 120 Hz (triangles), respectively. $I_{WB,ext} = 3.27 \text{ W cm}^{-2}$, $I_g = 2 \text{ W cm}^{-2}$, $E_{ext} = 60 \text{ V } \mu\text{m}^{-1}$. **b**, End values of the external diffraction efficiencies achieved at different hologram write rates with $I_{WB,ext} = 0.65$ (squares), 3.27 (circles) and 6.54 W cm^{-2} (triangles), respectively. $I_g = 2 \text{ W cm}^{-2}$, $E_{ext} = 60 \text{ V } \mu\text{m}^{-1}$. Solid lines are guides to the eye. The dashed line marks the video-rate frequency.

doping (oxidation) of the material, providing permanent traps in the material. This allowed us to achieve part of the sensitivity enhancement even without pre-illumination (E.M. and K.M., manuscript in preparation).

We checked whether the new gating effect is a unique property of the TPD-PPV-based material used here or whether it can be observed in composites based on the commonly used hole-conducting matrix poly(*N*-vinylcarbazole) (PVK) using the same azo dyes as electro-optical components. When sensitized with 2,4,7-trinitrofluorenone (TNF) or the C₆₀ derivative PCBM (see Methods) the gating effect was present. No gating effects were found in PVK-based materials sensitized with 2,4,7-trinitrofluorenonemalononitrile (TNFM), the sensitizer usually employed for carbazole-containing hole-conductors in the near-infrared region (E.M. and K.M., manuscript in preparation).

To demonstrate the high sensitivity of our new material, we performed a pulsed recording experiment. During the first three quarters of a recording cycle both write beams were on and the gate beam was off, and vice versa during the last quarter. Thus, the gate beam also serves efficiently to erase the previously written hologram. During each cycle a hologram with an external diffraction efficiency of several per cent can be written and be completely erased (Fig. 3a). This experiment was performed for different hologram write rates and different write beam intensities (Fig. 3b). At a write rate of 90 Hz and a write beam intensity of $I_{WB,ext} = 0.64 \text{ W cm}^{-2}$ ($I_{WB,int} = 0.2 \text{ W cm}^{-2}$), a value of $\eta_{ext} \approx 1\%$ was still achieved, yielding a sensitivity of $19 \text{ cm}^2 \text{ J}^{-1}$. No long-term changes of the PR performance under permanent use (overnight video-rate experiment corresponding to about 1.2 million completed exposure and erasure cycles) were observed.

Our devices are about one order of magnitude more sensitive at a given external field than the previously best organic PR devices: a multifunctional glass²³ and a methine dye²⁴ (Table 1). The latter were about twice as sensitive as the best PVK/TNFM-based systems^{25,26}. The PVK/TNFM-based composite investigated here (Table 1) had reduced sensitivity by even more than two orders of magnitude. We point out that a true comparison is difficult, because of differences in the experimental conditions employed by the different groups, such as device thickness, operating wavelength and the electric field strength.

Inorganic materials such as multiple quantum wells⁴ and Rh-doped BaTiO₃ crystals³ have been used for TGH in the near-infrared. Multiple quantum wells operate in the thin-grating limit ($d \approx 1 \mu\text{m}$; Raman–Nath regime), whereas the PR crystals are rather thick recording media (d is several mm; Bragg regime). Despite their relatively small active-layer thickness ($d \approx 100 \mu\text{m}$)

high-performance PR polymer devices show much higher diffraction efficiencies than the inorganic materials. Their response time is slower than in multiple quantum well devices ($\tau < 1 \text{ ms}$), but faster than in Rh-BaTiO₃ ($\tau \approx 1 \text{ s}$; ref. 3). Overall, the sensitivity (equation (1)) of organic PR devices is higher than in the inorganic counterparts. Furthermore, they offer the important advantage of a large aperture. This permits holographic recording with high spatial resolution (diffraction limited, that is, $3 \mu\text{m}$ in our case) and high angular bandwidth⁵.

The characterization presented so far was performed under ‘ideal’ (laboratory) conditions, that is, with plane-wave write beams, rather high intensity, and with high fringe contrast m (≈ 1). Under ‘real’ conditions much diffraction efficiency is lost because the object wave is distorted and usually much weaker than the reference wave, yielding poor contrast m . To demonstrate that our devices can still perform under these conditions, we set up a video-rate (30 Hz) imaging holographic system. Considering the low total intensity, the unfavourable write beam intensity ratio (~ 15) and the rather strong readout beam (yielding $m \approx 0.4$; see Methods for details), a respectable overall diffraction efficiency of the entire setup of 10^{-3} (at $E_{ext} = 60 \text{ V } \mu\text{m}^{-1}$) was nevertheless measured at the image output plane, allowing readout with a low-cost charge-coupled device (CCD) camera. In our case, the strong readout beam mimics the high background given by the diffusely scattered photons in a TGH experiment. In a real application readout would be performed with a short pulse at the end of the write phase and not continuously.

The degree of scattering as well as the required resolution in the biological specimen will eventually determine the depth up to which OCT and TGH can be used. Typical values range from 0.5 mm to 2 mm (ref. 6). If the ballistic photons are a fraction of 10^{-4} of the total photons (that is, 99.99% scatter) the contrast factor would be $m \approx 0.01$, a reduction of m by a factor of about 40 compared with our video-rate experiment (see above). According to equations (3) to (5) (see ‘parameter determination’ section) this would reduce the diffraction efficiency by a factor of 40^2 , yielding an overall $\eta \approx 10^{-6}$ assuming constant recording speed (E.M. and K.M., manuscript in preparation). The reduction in signal can be at least partially compensated for by using more sensitive CCD detectors.

In both OCT and TGH, the depth resolution (z -direction) is given by the bandwidth of the low-coherence light source^{6,7}. It is 10–15 μm in commercial OCT systems using normal super-luminescence diodes. Up to 1–3 μm resolution has been achieved by ultra-short laser pulses⁷. For OCT the lateral resolution (x – y plane) is given by the spacing between adjacent scan points, whereas TGH provides diffraction-limited x – y resolution (3 μm in the geometry

Table 1 Near-infrared sensitivity comparison of selected organic photorefractive materials

Material composition (wt%)	Wavelength of write beam, λ_{WB} (nm)	Intensity of write beam, I_{WB} (W cm^{-2})	External field, E_{ext} ($\text{V } \mu\text{m}^{-1}$)	Exposure time, t_{exp} (s)	Sensitivity, S ($\text{cm}^2 \text{ J}^{-1}$)	S rescaled to $60 \text{ V } \mu\text{m}^{-1}$ ($\text{cm}^2 \text{ J}^{-1}$)†	Ref.
AZO:TPD-PPV:DPP:PCBM 30:56:13:1, gated at 633 nm (2 W cm^{-2})	830	0.65	60	0.008	19	19	This work
DRDCTA:EHMPA:TNFM 69:30:1	790	1	80	~ 0.013	7.7	~ 3.2	23
Methine dye 100	780	3	84	$\sim 0.005^*$	6.7	~ 2.4	24
Chrom. C:PVK:ECZ:TNFM 28.5:45:25.5:1	780	5	48	0.03	0.67	~ 1.3	25
DHAC-MPN:PVK:DPP:TNFM 25:49:24:2	830	1	29	0.8	0.13	~ 1.1	26
AZO:PVK:ECZ:TNFM 30:46:23:1	830	3.27	60	0.37	0.08	0.08	This work

* Estimated from fit parameters given in ref. 24.

† The sensitivity is strongly field dependent. For comparing the devices we tried to rescale the sensitivities to a field to 60 V cm^{-1} by assuming $\sqrt{\eta_{ext}} \propto E_{ext}^2$ (typically observed in low- T_g materials^{12,13}) and $t_{exp} \propto E_{ext}^{-1}$. We point out that this procedure tends to underestimate (overestimate) S for upscaling (downscaling) because the field dependence of the recording time of real system is generally stronger than linear (for example, ref. 23). AZO, azo dye mixture described in Methods.

used here) owing to the parallel imaging character of holography. Thus, in TGHI the processing time is independent of the x - y image size, but in OCT this scales with the area and the lateral resolution, that is, with the total number of data points to be scanned. For comparison, OCT systems require about 10 s for recording one x - y image plane of $1.5 \times 2 \text{ mm}^2$ with $5 \mu\text{m}$ resolution²⁷. By using an exposure time longer than 33 ms for the material presented here we could regain diffraction efficiency and still achieve a more favourable recording rate in a TGHI system.

Whereas achieving high refractive index modulation amplitudes with optimized nonlinear-optical chromophores is well understood¹⁵, this work opens the way for achieving high recording speeds. Now, sufficiently large external diffraction efficiencies are feasible without sacrificing speed. Gating may make it possible to extend the operation wavelength of organic PR materials further towards the infrared.

We have demonstrated multiple-video-rate holographic storage with low light power and moderate field strength using a real object. Similar experiments using near-infrared laser light were not previously possible. The availability of low-cost, high-power NIR laser diodes and sensitive CCD chips make the development of an operating time-gated holographic imaging system seem economically feasible. TGHI promises faster recording time than commercial OCT, particularly for large regions of interest. One benefit would be a reduced tendency towards 'motion artefacts' which are often anticipated in *in vivo* studies by OCT. □

Methods

Material composition

The material investigated was based on TPD-PPV (56 wt%, chemical structure shown in Fig. 1b) as the photoconductive host matrix. The eutectic mixture of two azo dyes (2,5-dimethyl-(4-*p*-nitrophenylazo)-anisole and 3-methoxy-(4-*p*-nitrophenylazo)-anisole, ratio 1:1; 30 wt%) was used as the electro-optical component. Diphenyl-phthalate (DPP, 13 wt%) was used as a plasticizer to ensure sufficient orientational mobility for the chromophores. Finally, as the sensitizer, 1 wt% of the highly soluble C₆₀ derivative [6,6]-phenyl-C₆₁-butyric acid-methylester (PCBM) was added²⁸. The resulting composite had a glass transition temperature $T_g \approx 10^\circ\text{C}$ (differential scanning calorimetry, heating rate + 20 K min⁻¹); measurements were made at 22 °C. The absorption coefficients at 830 (633) nm were 5 (155) cm⁻¹.

Holographic set-up parameters

Holographic experiments were performed (40-mW laser diode) with two s-polarized write beams (1 and 2) with external angles $\alpha_{1,\text{ext}} = 50.8^\circ$ and $\alpha_{2,\text{ext}} = 71.1^\circ$ relative to the sample normal. The write beams were overlapped in 105- μm -thick polymer films sandwiched between ITO/glass electrodes. The full-width at half-maximum (FWHM) of the plane gaussian writing beams was 0.47 mm. Owing to the tilted geometry, the illuminated area is elliptical with a half-maximum surface of 0.273 mm² (0.535 mm²) for write beam 1 (2). Reflection losses in the multilayer device before entering the composite were calculated to be 13% (33%) for write beam 1 (2). 1 W cm⁻² internal intensity corresponds to an external intensity of 3.27 W cm⁻². The external diffraction efficiency η_{ext} was determined in degenerate four-wave mixing experiments. Readout was performed by a weak p-polarized probe beam (external intensity $I_{R,\text{ext}}$, diffracted component $I_{R,\text{diff}}$) counter-propagating to beam 1.

Parameter determination

η_{ext} was calculated according to

$$\eta_{\text{ext}} = I_{R,\text{diff}} / I_{R,\text{ext}} \quad (2)$$

η_{ext} depends on the refractive index modulation amplitude Δn according to

$$\eta_{\text{ext}} = R \exp(-\alpha d / \cos \alpha_1) \eta_{\text{int}} \quad (3a)$$

$$\eta_{\text{int}} = \sin^2 \left(\frac{\pi \Delta n d}{\lambda \cos \alpha_1} \right) \quad (3b)$$

Here η_{int} is the internal diffraction efficiency, α is the absorption coefficient, d is the sample thickness, λ is the laser wavelength, α_1 is the internal angle of the read beam relative to the sample normal, and $R < 1$ is a factor taking into account reflection losses. Δn is given by

$$\Delta n \approx E_0 E_{\text{SC}} \epsilon_{\text{EO,eff}} \quad (4)$$

$\epsilon_{\text{EO,eff}}$ is the effective electro-optical coefficient, which by itself depends on number density of EO chromophores and the PR molecular figure-of-merit (FOM) of the chromophore¹³. E_{sc} is the space-charge field, which is proportional¹¹ to the contrast factor m of the grating given by

$$m = 2(I_1 I_2)^{1/2} / I_{\text{tot}} \quad (5)$$

where I_{tot} is the total light intensity incident on the device, including read beam and incoherent illumination (scattered light in TGHI). For the systematic studies the external write-beam powers were adjusted such that the internal intensities were equal, that is, $m \approx 1$ owing to the weak read beam intensity.

The field-dependent two-beam coupling gain coefficient $\Gamma(E)$, which describes the energy transfer between the write beams, was calculated from the transmitted write-beam intensities I_1 and I_2 according to

$$\Gamma = \frac{1}{d} \left[\cos \alpha_1 \ln \frac{I_1(E)}{I_1(E=0)} - \cos \alpha_2 \ln \frac{I_2(E)}{I_2(E=0)} \right] \quad (6)$$

According to ref. 9 Γ can be approximated by

$$\Gamma \approx C_{\text{TBC}} \Delta n \sin(\varphi_g) \quad (7)$$

with $C_{\text{TBC}} = 2\pi / [\lambda_0 \cos[(\alpha_2 - \alpha_1)/2]]$, φ_g is the phase shift between the light interference pattern and the PR index grating. The trap density N_T was calculated from Δn and Γ according to the standard Kukhtarev model developed for inorganic PR crystals¹¹.

As a metric of comparison and to qualitatively visualize the influence that the gating has on the recording dynamics, we use the time to reach 50% of the quasi-steady-state diffraction efficiency after 60 s of recording, τ_{50} .

Recording scheme

The recording dynamics without gating was determined by switching both write beams on after a pre-poling period in the dark $t_p = 300$ s (much longer than the relaxation time; see main text) and monitoring of $I_{R,\text{diff}}$. For the gating experiments a gate pulse from a 633-nm HeNe-laser with duration t_g and intensity I_g was applied at normal incidence during the pre-poling period right before switching on the near-infrared write beams.

Holographic imaging set-up

The video-rate holographic imaging set-up contained a moving slide ($1 \times 2 \text{ mm}^2$ object size) as the input mask. Read-out was performed with the back-reflected reference wave after rotating its polarization by 90° to p-polarization. The total external powers of the 830 nm write and read beams (FWHM, 1.3 mm) were $P_{\text{obj}} = 0.4 \text{ mW}$, $P_{\text{ref}} = 5.8 \text{ mW}$, $P_{\text{read}} = 1.8 \text{ mW}$; gating was performed at 633 nm with $P_{\text{gate}} = 8 \text{ mW}$ (FWHM, 1.3 mm). According to equation (5) the average grating contrast was $m \approx 0.4$.

Received 31 October 2001; accepted 8 July 2002; doi:10.1038/nature00975.

- Burr, G. W. & Leyva, I. Multiplexed phase-conjugate holographic data storage with a buffer hologram. *Opt. Lett.* **25**, 499–501 (2000).
- Stepanov, S. in *Handbook of Advanced Electronic and Photonic Materials and Devices* Ch. 6 (ed. Nalva, H. S.) Vol. 2 (Academic, New York, 2001).
- Hyde, S. C. W. *et al.* Depth-resolved holographic imaging through scattering media by photorefractive. *Opt. Lett.* **20**, 1331–1334 (1995).
- Jones, R. *et al.* Holographic storage and high background imaging using photorefractive multiple quantum wells. *Appl. Phys. Lett.* **69**, 1837–1839 (1996).
- Steele, D. D. *et al.* Transillumination imaging through scattering media by use of photorefractive polymers. *Opt. Lett.* **23**, 153–155 (1998).
- Huang, D. *et al.* Optical coherence tomography. *Science* **254**, 1178–1181 (1991).
- Drexler, W. *et al.* *In vivo* ultrahigh-resolution optical coherence tomography. *Opt. Lett.* **24**, 1221–1223 (1999).
- Hörhold, H. H. *et al.* Synthesis of TPD-containing polymers for use as light-emitting materials in electroluminescent and laser devices. *Proc. SPIE* **4105**, 431–442 (2001).
- Moerner, W. E., Grunnet-Jepsen, A. & Thompson, C. L. Photorefractive polymers. *Annu. Rev. Mater. Sci.* **27**, 585–623 (1997).
- Zilker, S. Materials design and physics of organic photorefractive systems. *ChemPhysChem* **1**, 72–87 (2000).
- Kukhtarev, N. V., Markov, V. B., Odulov, S. G., Soskin, M. S. & Vinetskii, V. L. Holographic storage in electrooptic crystals. I. Steady state. *Ferroelectrics* **22**, 949–960 (1979).
- Moerner, W. E., Silence, S. M., Hache, F. & Björklund, G. C. Orientationally enhanced photorefractive effect in polymers. *J. Opt. Soc. Am.* **11**, 320–330 (1994).
- Wortmann, R. *et al.* Design of optimized photorefractive polymers: A novel class of chromophores. *J. Chem. Phys.* **105**, 10637–10647 (1996).
- Ashley, J. *et al.* Holographic data storage. *IBM J. Res. Dev.* **44**, 341–367 (2000).
- Würthner, F., Wortmann, R. & Meerholz, K. Chromophore design for photorefractive organic materials. *ChemPhysChem* **3**, 17–31 (2002).
- Silence, S. M., Björklund, G. C. & Moerner, W. E. Optical trap activation in a photorefractive polymer. *Opt. Lett.* **19**, 1822–1824 (1994).
- Herlocker, J. A. *et al.* Stabilization of the response time in photorefractive polymers. *Appl. Phys. Lett.* **77**, 2292–2294 (2000).
- Wolff, J., Schlöter, S., Hofmann, U., Haarer, D. & Zilker, S. Speed enhancement of photorefractive polymers by means of light-induced filling of trapping states. *J. Opt. Soc. Am. B* **16**, 1080–1086 (1999).
- Buse, K., Adibi, A. & Psaltis, D. Non-volatile holographic storage in doubly doped lithium niobate crystals. *Nature* **393**, 665–668 (1998).
- Günther, H., Macfarlane, R., Furukawa, Y., Kitamura, K. & Neurgaonkar, R. Two-color holography in reduced near-stoichiometric lithium niobate. *Appl. Opt.* **37**, 7611–7623 (1998).
- Grunnet-Jepsen, A. Spectroscopic determination of trap density in C60-sensitized photorefractive polymers. *Chem. Phys. Lett.* **291**, 553–561 (1998).
- Wang, L., Ng, M.-K. & Yu, L. Photorefractive and complementary grating competition in bipolar transport molecular material. *Phys. Rev. B* **62**, 4973–4984 (2000).
- Zilker, S. J. & Hofmann, U. Organic photorefractive glass with infrared sensitivity and fast response. *Appl. Opt.* **39**, 2287–2290 (2000).
- Wang, L., Ng, M. K. & Yu, L. Efficient molecular photorefractive materials based on methine dyes. *Appl. Phys. Lett.* **78**, 700–702 (2001).
- Steenwinckel, D. V., Hendrickx, E., Persoons, A., Van den Broeck, K. & Samyn, C. Influence of the

- chromophore ionization potential on speed and magnitude of photorefractive effects in poly-N-vinylcarbazole based polymer composites. *J. Chem. Phys.* **112**, 11030–11037 (2000).
26. Kippelen, B. *et al.* Infrared photorefractive polymers and their applications for imaging. *Science* **279**, 54–57 (1998).
27. Boppart, S. A. *et al.* In vivo cellular optical coherence tomography imaging. *Nature Med.* **4**, 861–865 (1998).
28. Hummelen, J. C. *et al.* Preparation and characterization of fulleroid and methanofullerene derivatives. *J. Org. Chem.* **60**, 532–538 (1995).

Acknowledgements

We thank R. Bittner, D. Müller, M. Hofmann and R. Birngruber for discussions. Financial support was granted by the Volkswagen Foundation (Germany), the European Space Agency (MAP-project), the Fonds der Chemischen Industrie (Germany), and the Bavarian government through 'Neue Werkstoffe' (Germany).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.M.
(e-mail: klaus.meerholz@uni-koeln.de).

Hydrogen from catalytic reforming of biomass-derived hydrocarbons in liquid water

R. D. Cortright, R. R. Davda & J. A. Dumesic

Department of Chemical Engineering, University of Wisconsin, Madison, Wisconsin 53706, USA

Concerns about the depletion of fossil fuel reserves and the pollution caused by continuously increasing energy demands make hydrogen an attractive alternative energy source. Hydrogen is currently derived from nonrenewable natural gas and petroleum¹, but could in principle be generated from renewable resources such as biomass or water. However, efficient hydrogen production from water remains difficult and technologies for generating hydrogen from biomass, such as enzymatic decomposition of sugars^{2–5}, steam-reforming of bio-oils^{6–8} and gasification⁹, suffer from low hydrogen production rates and/or complex processing requirements. Here we demonstrate that hydrogen can be produced from sugars and alcohols at temperatures near 500 K in a single-reactor aqueous-phase reforming process using a platinum-based catalyst. We are able to convert glucose—which makes up the major energy reserves in plants and animals—to hydrogen and gaseous alkanes, with hydrogen constituting 50% of the products. We find that the selectivity for hydrogen production increases when we use molecules that are more reduced than sugars, with ethylene glycol and methanol being almost completely converted into hydrogen and carbon dioxide. These findings suggest that catalytic aqueous-phase reforming might prove useful for the generation of hydrogen-rich fuel gas from carbohydrates extracted from renewable biomass and biomass waste streams.

We consider production of hydrogen by low-temperature reforming (at 500 K) of oxygenated hydrocarbons having a C:O stoichiometry of 1:1. For example, reforming of the sugar-alcohol sorbitol to H₂ and CO₂ occurs according to the following stoichiometric reaction:



The equilibrium constant for reaction (1) per mole of CO₂ is of the order of 10⁸ at 500 K, indicating that the conversion of sorbitol in

the presence of water to H₂ and CO₂ is highly favourable. However, the selective generation of hydrogen by this route is difficult because the products H₂ and CO₂ readily react at low temperatures to form alkanes and water. For example, the equilibrium constant at 500 K for the conversion of CO₂ and H₂ to methane (reaction 2) is of the order of 10¹⁰ per mole of CO₂.

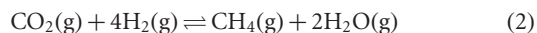


Figure 1 shows a schematic representation of the reaction pathways we believe to be involved in the formation of H₂ and alkanes from oxygenated hydrocarbons over a metal catalyst. The reactant undergoes dehydrogenation steps on the metal surface to give adsorbed intermediates before the cleavage of C–C or C–O bonds. With platinum, the catalyst we use, the activation energy barriers for cleavage of O–H and C–H bonds are similar¹⁰; however, Pt–C bonds are more stable than Pt–O bonds, so adsorbed species are probably bonded preferentially to the catalyst surface through Pt–C bonds. Subsequent cleavage of C–C bonds leads to the formation of CO and H₂, and CO reacts with water to form CO₂ and H₂ by the water–gas shift reaction (that is, CO + H₂O ⇌ CO₂ + H₂)^{11,12}.

The further reaction of CO and/or CO₂ with H₂ leads to alkanes and water by methanation and Fischer–Tropsch reactions^{13–15}; this H₂ consuming reaction thus represents a series-selectivity challenge. In addition, undesirable alkanes can form on the catalyst surface by cleavage of C–O bonds, followed by hydrogenation of the resulting adsorbed species. This process constitutes a parallel-selectivity challenge. Another pathway that contributes to this parallel-selectivity challenge is cleavage of C–O bonds through dehydration reactions catalysed by acidic sites associated with the catalyst support^{16,17} or catalysed by protons in the aqueous solution^{18,19}, followed by hydrogenation reactions on the catalyst. In addition, organic acids can be formed by dehydrogenation reactions catalysed by the metal, followed by rearrangement reactions²⁰ that take place in solution or on the catalyst. These organic acids lead to the formation of alkanes from carbon atoms that are not bonded to oxygen atoms.

Table 1 summarizes our experimental results for aqueous-phase reforming of glucose, the compound most relevant to hydrogen production from biomass, as well as for the reforming of sorbitol, glycerol, ethylene glycol and methanol. Reactions were carried out over a Pt/Al₂O₃ catalyst at 498 and 538 K (see Methods for experimental details). The fractions of the feed carbon detected in the effluent gas and liquid streams yield a complete carbon balance for all feed molecules, indicating that negligible amounts of carbon have been deposited on the catalyst. Catalyst performance was stable

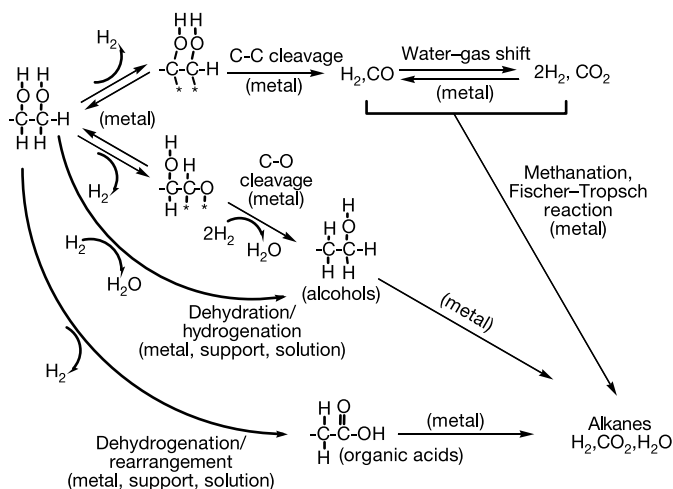


Figure 1 Reaction pathways for production of H₂ by reactions of oxygenated hydrocarbons with water. (Asterisk represents a surface metal site.)

for long periods of time on stream (for example, 1 week). Results from replicate runs agree to within $\pm 3\%$.

The hydrogen selectivities shown in Table 1 are defined as the number of H_2 molecules detected in the effluent gas, normalized by the number of H_2 molecules that would be present if the carbon atoms detected in the effluent gas molecules had all participated in the reforming reaction. (That is, we infer the amount of converted glucose, sorbitol, glycerol, ethylene glycol and methanol from the carbon-containing gas-phase products and assume that each of the feed molecules would yield 2, 13/6, 7/3, 5/2 or 3 molecules of H_2 , respectively.) Alkane selectivity is defined as the number of carbon atoms in the gaseous alkane products normalized by the total number of carbon atoms in the gaseous effluent stream.

Figure 2 illustrates that the selectivity for H_2 production improves in the order glucose < sorbitol < glycerol < ethylene glycol < methanol. Figure 2 also implies that lower operating temperatures result in higher H_2 selectivities, although this trend is in part due to the lower conversions achieved at lower temperatures. The selectivity for alkane production follows a trend with respect to reactant that is opposite to that exhibited by the H_2 selectivity.

The gas streams from aqueous-phase reforming of the oxygenated hydrocarbons were found to contain low levels of CO (that is, less than 300 p.p.m.). For aqueous-phase reforming of glycerol, where analysis of liquid-phase products is more tractable compared to sorbitol and glucose, the major reaction intermediates detected include (in approximate order of decreasing concentration) ethanol, 1,2-propanediol, methanol, 1-propanol, acetic acid, ethylene glycol, acetol, 2-propanol, propionic acid, acetone, propionaldehyde, and lactic acid. Analysis of the gas phase effluent indicates the presence of trace amounts of methanol and ethanol (about 300 p.p.m.).

High hydrogen yields are only obtained when using sorbitol, glycerol and ethylene glycol as feed molecules for aqueous-phase reforming. Although these molecules can be derived from renewable feedstocks^{21–24}, the reforming of less reduced and more immediately available compounds such as glucose is likely to be more practical; but hydrogen yields for glucose reforming are relatively low (Table 1 and Fig. 2). However, we expect that improvements in catalyst performance, reactor design, and reaction conditions may increase the hydrogen selectivity for the direct aqueous-phase reforming of sugars. For example, the lower H_2 selectivities for the aqueous-phase reforming of glucose, compared to that achieved using the other oxygenated hydrocarbon reactants, are caused at least partially by homogeneous reactions of glucose in the aqueous phase at the temperatures used in this study¹⁹. Thus, higher selectivities for H_2 from aqueous-phase reforming of glucose might be achieved by reactor designs that maximize the number of catalytically active sites (leading to desirable surface reactions) and

minimize the void volume (leading to undesirable liquid-phase reactions) in the reactor.

Here, using Pt/Al_2O_3 as catalyst, we found that lower glucose concentrations correlated with higher selectivities for hydrogen production, and reforming experiments were therefore conducted at low feed concentrations of 1 wt%, which corresponds to a molar ratio H_2O/C_1 of 165. Processing such dilute solutions is economically not practical, even though reasonably high hydrogen yields are achieved (Table 1). However, undesirable homogeneous reactions, as observed with glucose, pose less of a problem when using sorbitol, glycerol, ethylene glycol and methanol, which makes it possible to generate high yields of hydrogen by the aqueous-phase reforming of more concentrated solutions containing these compounds. For example, we have found that upon increasing the feed concentrations of ethylene glycol or glycerol to 10 wt% (molar ratio $H_2O/C_1 = 15$), it is still possible to achieve high conversions and high selectivities for H_2 production. (A hydrogen selectivity of 97% was achieved with 62% conversion of 10 wt% ethylene glycol; and a hydrogen selectivity of 70% was achieved with 77% conversion of 10 wt% glycerol.) A molar H_2O/C_1 ratio of 15 is still higher than that typically utilized in conventional vapour-phase reforming of hydrocarbons (molar $H_2O/C_1 = 3$ to 5); but our aqueous-phase reforming system has the advantage of not requiring energy-intensive vaporization of water to generate steam.

Operating at higher reactant concentrations and lower conversion levels leads to higher rates of hydrogen production. Rates of hydrogen production are measured as turnover frequencies (that is, rates normalized by the number of surface metal atoms as determined from the irreversible uptake of CO at 300 K). For example, the turnover frequency for production of hydrogen at 498 K from a 1 wt% ethylene glycol solution is 0.08 min^{-1} for the 90% conversion run listed in Table 1 (weight-hourly space velocity, $WHSV = 0.008 \text{ g of ethylene glycol per g of catalyst per h}$), but increases to 0.7 min^{-1} for a feed containing 10 wt% ethylene glycol (run at 498 K and at a higher $WHSV$ of 0.12 g of ethylene glycol per g of catalyst per h). The hydrogen selectivity for this latter run is equal to 97% at a conversion of 62%. The turnover frequency for hydrogen production from 10 wt% ethylene glycol at 498 K increased further to 7 min^{-1} when higher space-velocities were used ($WHSV = 18 \text{ g of ethylene glycol per g of catalyst per h}$) with a highly dispersed 3 wt% Pt/Al_2O_3 catalyst consisting of smaller alumina particles (63–125 μm) to minimize transport limitations. The hydrogen selectivity for this kinetically controlled run was 99% at a conversion of 3.5%.

The rates of formation of H_2 from aqueous-phase reforming of 10 wt% glucose, sorbitol, glycerol, ethylene glycol and methanol at 498 K over a 3 wt% Pt/Al_2O_3 catalyst under conditions to minimize transport limitations and at low conversions of the reactant have

Table 1 Experimental data for reforming of oxygenated hydrocarbons

	Glucose		Sorbitol		Glycerol		Ethylene glycol		Methanol	
Temperature (K)	498	538	498	538	498	538	498	538	498	538
Pressure (bar)	29	56	29	56	29	56	29	56	29	56
% Carbon in liquid-phase effluent	51	15	39	12	17	2.8	11	2.9	6.5	6.4
% Carbon in gas-phase effluent	50	84	61	90	83	99	90	99	94	94
Gas-phase composition										
H_2 (mol. %)	51	46	61	54	64.8	57	70	68.7	74.6	74.8
CO_2 (mol. %)	43	42	35	36	29.7	32	29.1	29	25	24.6
CH_4 (mol. %)	4.0	7.0	2.5	6.0	4.2	8.3	0.8	2.0	0.4	0.6
C_2H_6 (mol. %)	2.0	2.7	0.7	2.3	0.9	2.0	0.1	0.3	0.0	0.0
C_3H_8 (mol. %)	0.0	1.0	0.8	1.0	0.4	0.7	0.0	0.0	0.0	0.0
C_4 , C_5 , C_6 alkanes (mol. %)	0.0	1.2	0.0	0.6	0.0	0.0	0.0	0.0	0.0	0.0
% H_2 selectivity*	50	36	66	46	75	51	96	88	99	99
% Alkane selectivity†	14	33	15	32	19	31	4	8	1.7	2.7

The catalyst was loaded in a tubular reactor, housed in a furnace, and reduced prior to reaction kinetics studies. The reactor system was pressurized with N_2 , and the reforming reaction was carried out at the listed reaction conditions. Each reaction condition was run for 24 h, during which the experimental data were collected. Further experimental details are provided in the Methods.

*% H_2 selectivity = (molecules H_2 produced/C atoms in gas phase)(1/RR) $\times 100$, where RR is the H_2/CO_2 reforming ratio, which depends on the reactant compound. RR values for the compounds are: glucose, 2; sorbitol, 13/6; glycerol, 7/3; ethylene glycol, 5/2; methanol, 3. We note that H_2 and alkane selectivities do not add up to 100%, because they are based independently on H-balances and C-balances, respectively. % Carbon in gas and liquid phase effluents add to 100% for a complete carbon balance. Slight $\pm$ deviations from 100% are caused by experimental error.

†% Alkane selectivity = (C atoms in gaseous alkanes/total C atoms in gas phase product) $\times 100$.

been measured to be 0.5, 1.0, 3.5, 7.0 and 7.0 min⁻¹, respectively. These turnover frequencies correspond to hydrogen production rates of 50, 100, 350, 700, and 700 l of H₂ (at standard temperature and pressure, 273 K and 1.01 bar) per l of reactor volume per h for each feed molecule, respectively. These rates compare favourably to the maximum rate of hydrogen production from glucose of about 5 × 10⁻³ l of H₂ per l of reactor volume per hour by enzymatic routes⁵. Normalization of the rates by the mass of catalyst used yields rates of about 3 × 10³, 6 × 10³, 2 × 10⁴, 4 × 10⁴, and 4 × 10⁴ μmol g⁻¹ h⁻¹ for hydrogen production at 498 K, from 10 wt% glucose, sorbitol, glycerol, ethylene glycol and methanol, respectively. These rates can be compared to the maximum value of 7 × 10² μmol g⁻¹ h⁻¹ reported for hydrogen production from glucose by enzymatic routes. (For this comparison, we have assumed a typical value of 100 units per mg of protein, where a unit of enzyme activity corresponds to the amount of enzyme which under standard assay conditions converts 1 μmol of substrate per min).

If the hydrogen produced were fed to a fuel cell operating at 50% efficiency, the rate of hydrogen production in our reformer would generate approximately 1 kW of power per l of reactor volume. Electrical power might thus be generated cost-effectively by an integrated fuel-cell/liquid-phase reformer system fed with a low-cost carbohydrate stream derived from waste biomass (for example, corn stover, wheat straw, wood waste). The practical use of aqueous-phase reforming reactions in this manner would depend on efficient feed recycling strategies and on efficient separation of hydrogen from the gaseous effluent stream. The gaseous effluents separated from the main product hydrogen could be combusted to generate the energy necessary for the liquid-phase-reforming reactor. However, some fuel cell applications might not require extensive purification because the main components in the reformer gas effluent other than hydrogen are CO₂ and methane, which can act as diluents²⁵. Alcohols and organic acids are also present in the gas effluent, but only at trace levels (300 p.p.m. and about 5 p.p.m., respectively) which may not lead to irreversible poisoning²⁶.

Reforming reactions between hydrocarbons and water to generate hydrogen are endothermic, and conventional steam-reforming of petroleum thus depends on the combustion of additional hydrocarbons to provide the heat needed to drive the reforming reaction. In contrast, the energy required for the aqueous-phase reforming of oxygenated hydrocarbons may be produced internally, by allowing a fraction of the oxygenated compound to form alkanes through exothermic reaction pathways. In this respect, the

formation of a mixture of hydrogen and alkanes from aqueous-phase reforming of glucose, as accomplished in the present study, is essentially neutral energetically, and little additional energy is required to drive the reaction. In fact, the energy contained in these alkanes could be used as a feed to an internal combustion engine or suitable fuel cell; this would allow the use of biomass-derived energy to drive the aqueous-phase reforming of glucose (and biomass more generally) with high yields to renewable energy.

While the present findings establish that Pt-based catalysts show high activities and good selectivity for the production of hydrogen from sugars and alcohols by aqueous-phase reforming reactions, improvements are necessary to render the process useful. Highly active catalytic materials that can satisfy the series and parallel selectivity challenges outlined in Fig. 1, but at a lower materials cost than for Pt, are particularly desirable. Moreover, new combinations of catalysts and reactor configurations are needed to obtain higher hydrogen yields from more concentrated solutions of glucose, given that glucose is the only compound we have tested that is directly relevant to biomass utilization. We believe that such improvements are possible, for example, by searching for catalysts that exhibit higher activity at lower temperatures, to minimize the deleterious effects of homogeneous decomposition reactions. □

Methods

Experiments for the aqueous-phase reforming of glucose, sorbitol, glycerol, ethylene glycol and methanol were performed over a 3 wt% Pt catalyst supported on nanofibres of γ-alumina (500 m² g⁻¹, Argonide Corp.). The catalyst was prepared by incipient wetness impregnation of alumina with tetraamine platinum nitrate solution, followed by drying at 380 K, calcination at 533 K in flowing oxygen and reduction at 533 K in flowing hydrogen. Chemisorption experiments using carbon monoxide at 300 K showed a CO uptake of 105 μmol per g of catalyst. A stainless steel tubular reactor (having an inner diameter of 5 mm and length of 45 cm) was loaded with 4.5 g of the pelletized Pt/Al₂O₃ catalyst, which was then reduced under flowing hydrogen at 533 K. The total pressure of the system was then increased by addition of nitrogen to a value slightly higher than the vapour pressure of water at the reaction temperature. The system pressure was controlled by a backpressure regulator. An aqueous solution containing 1 wt% of the oxygenated compound was fed continuously, using a high-performance liquid chromatography (HPLC) pump, at 3.6 ml h⁻¹ into the reactor heated to the desired reaction temperature. Under these conditions, the WHSV was 0.008 g of oxygenated compound per g of catalyst per h through the reactor. The effluent from the reactor was water-cooled in a double-pipe heat exchanger to liquefy the condensable vapours. The fluid from this cooler was combined with the nitrogen make-up gas at the top of the cooler, and the gas and liquid were separated in a stainless-steel vessel (about 130 cm³) maintained at the system pressure. The effluent liquid was drained periodically (every 12 h) for total organic carbon (TOC) analysis and for detection of the primary carbonaceous species using gas chromatography and HPLC. The effluent gas stream passed through the back-pressure regulator and was analysed with several online gas chromatographs. The kinetic data for each reaction condition were typically collected over a 24-h period, after which the reaction conditions were changed. The catalyst performance was stable for times on stream of at least 1 week.

Received 6 February; accepted 23 July 2002; doi:10.1038/nature01009.

1. Rostrup-Nielsen, J. Conversion of hydrocarbons and alcohols for fuel cells. *Phys. Chem. Chem. Phys.* **3**, 283–288 (2001).
2. Kumar, N. & Das, D. Enhancement of hydrogen production by enterobacter cloacae IIT-BT-08. *Process Biochem.* **35**, 589–593 (2000).
3. Woodward, J. *et al.* Enzymatic hydrogen production: Conversion of renewable resources for energy production. *Energy Fuels* **14**, 197–201 (2000).
4. Yokoi, H. *et al.* Microbial hydrogen production from sweet potato starch residue. *J. Biosci. Bioeng.* **91**, 58–63 (2001).
5. Woodward, J., Orr, M., Cordray, K. & Greenbaum, E. Enzymatic production of biohydrogen. *Nature* **405**, 1014 (2000).
6. Garcia, L., French, R., Czernik, S. & Chornet, E. Catalytic steam reforming of bio-oils for the production of hydrogen: Effects of catalyst composition. *Appl. Catal. A* **201**, 225–239 (2000).
7. Amphlett, J. C., Leclerc, S., Mann, R. F., Peppley, B. A. & Roberge, P. R. Fuel cell hydrogen production by catalytic ethanol-steam reforming. *Proc. 33rd Intersoc. Energy Convers. Eng. Conf.* **269**, 1–7 (1998).
8. Markevich, M., Czernik, S., Chornet, E. & Montane, D. Hydrogen from biomass: Steam reforming of model compounds of fast-pyrolysis oil. *Energy Fuels* **13**, 1160–1166 (1999).
9. Milne, T. A., Elam, C. & Evans, R. J. *Hydrogen from Biomass: State of the Art and Research Challenges* 1–82 (National Renewable Energy Laboratory, Golden, CO, 2002).
10. Greeley, J. & Mavrikakis, M. A first-principles study of methanol decomposition on Pt(111). *J. Am. Chem. Soc.* **124**, 7193–7201 (2002).
11. Grenoble, D. C., Estadt, M. M. & Ollis, D. F. The chemistry and catalysis of the water gas shift reaction. 1. The kinetics over supported metal catalysts. *J. Catal.* **67**, 90–102 (1981).
12. Hilaire, S., Wang, X., Luo, T., Gorte, R. J. & Wagner, J. A comparative study of water-gas shift reaction over ceria supported metallic catalysts. *Appl. Catal. A* **215**, 271–278 (2001).
13. Iglesia, E., Soled, S. L. & Fiato, R. A. Fischer-Tropsch synthesis on cobalt and ruthenium. *Metal*

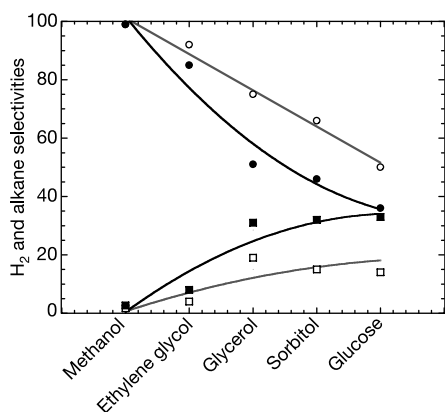


Figure 2 Selectivities (%) versus oxygenated hydrocarbon. H₂ selectivity (circles) and alkane selectivity (squares) from aqueous-phase reforming of 1 wt% oxygenated hydrocarbons over 3 wt% Pt/Al₂O₃ at 498 K (open symbols) and 538 K (filled symbols). The aqueous feed solution was fed to the reactor at a weight-hourly space velocity of 0.008 g of oxygenated hydrocarbon per gram of catalyst per hour. High conversions of the reactant were achieved under these conditions (50–99% conversion to gas-phase carbon, as indicated in Table 1) to provide a rigorous test of the carbon mass balance.

- dispersion and support effects on reaction rate and selectivity. *J. Catal.* **137**, 212–224 (1992).
14. Kellner, C. S. & Bell, A. T. The kinetics and mechanism of carbon monoxide hydrogenation over alumina-supported ruthenium. *J. Catal.* **70**, 418–432 (1981).
 15. Vannice, M. A. The catalytic synthesis of hydrocarbons from H₂/CO mixtures over the group VIII metals V. The catalytic behaviour of silica-supported metals. *J. Catal.* **50**, 228–236 (1977).
 16. Bates, S. P. & Van Santen, R. A. Molecular basis of zeolite catalysis: A review of theoretical simulations. *Adv. Catal.* **42**, 1–114 (1998).
 17. Gates, B. *Catalytic Chemistry* (Wiley, New York, 1992).
 18. Eggleston, G. & Vercellotti, J. R. Degradation of sucrose, glucose and fructose in concentrated aqueous solutions under constant pH conditions at elevated temperature. *J. Carbohydr. Chem.* **19**, 1305–1318 (2000).
 19. Kabyemela, B. M., Adschiri, T., Malaluan, R. M. & Arai, K. Glucose and fructose decomposition in subcritical and supercritical water: Detailed reaction pathway, mechanisms, and kinetics. *Ind. Eng. Chem. Res.* **38**, 2888–2895 (1999).
 20. Collins, P. & Ferrier, R. *Monosaccharides: Their Chemistry and Their Roles in Natural Products* (Wiley, West Sussex, England, 1995).
 21. Li, H., Wang, W. & Deng, J. F. Glucose hydrogenation to sorbitol over a skeletal Ni-P amorphous alloy catalyst (Raney Ni-P). *J. Catal.* **191**, 257–260 (2000).
 22. Blanc, B., Bourrel, A., Gallezot, P., Haas, T. & Taylor, P. Starch-derived polyols for polymer technologies: Preparation by hydrogenolysis on metal catalysts. *Green Chem.* **2**, 89–91 (2000).
 23. Narayan, R., Durrence, G. & Tsao, G. T. Ethylene glycol and other monomeric polyols from biomass. *Biotechnol. Bioeng. Symp.* **14**, 563–571 (1984).
 24. Tronconi, E. et al. A mathematical model for the catalytic hydrogenolysis of carbohydrates. *Chem. Eng. Sci.* **47**, 2451–2456 (1992).
 25. Larminie, J. & Dicks, A. *Fuel Cell Systems Explained* 189 (Wiley, West Sussex, England, 2000).
 26. Amphlett, J. C., Mann, R. F. & Peppley, B. A. On board hydrogen purification for steam reformer/PEM fuel cell vehicle power plants. *Hydrogen Energy Prog. X, Proc. 10th World Hydrogen Energy Conf.* **3**, 1681–1690 (1998).

Acknowledgements

We thank K. Allen, J. Shabaker and G. Huber for assistance in reaction kinetics measurements. We also thank G. Huber for help with catalyst preparation/characterization and for TOC analyses, and M. Sanchez-Castillo for assistance with analysis of reaction products. We thank M. Mavrikakis and researchers at Haldor Topsøe A/S for reviews and discussion. This work was supported by the US Department of Energy (DOE), Office of Basic Energy Sciences, Chemical Science Division.

Competing interests statement

The authors declare competing financial interests: details accompany the paper on Nature's website (<http://www.nature.com/nature>).

Correspondence and requests for materials should be addressed to J.A.D. (e-mail: dumesic@engr.wisc.edu).

Pfiesteria shumwayae kills fish by micropredation not exotoxin secretion

Wolfgang K. Vogelbein*, Vincent J. Lovko†, Jeffrey D. Shields†, Kimberly S. Reece*, Patrice L. Mason*, Leonard W. Haas* & Calvin C. Walker‡

* Virginia Institute of Marine Science, The College of William and Mary, Gloucester Point, Virginia 23062, USA

‡ United States Environmental Protection Agency, National Health and Environmental Effects Research Laboratory, Gulf Ecology Division, 1 Sabine Island Drive, Gulf Breeze, Florida 32561, USA

† These authors contributed equally to this work

Pfiesteria piscicida and *P. shumwayae* reportedly secrete potent exotoxins thought to cause fish lesion events, acute fish kills and human disease in mid-Atlantic USA estuaries^{1–7}. However, *Pfiesteria* toxins have never been isolated or characterized⁸. We investigated mechanisms by which *P. shumwayae* kills fish using three different approaches. Here we show that larval fish bioassays conducted in tissue culture plates fitted with polycarbonate membrane inserts exhibited mortality (100%) only in treatments where fish and dinospores were in physical contact. No mortalities occurred in treatments where the membrane prevented contact between dinospores and fish. Using differen-

tial centrifugation and filtration of water from a fish-killing culture, we produced 'dinoflagellate', 'bacteria' and 'cell-free' fractions. Larval fish bioassays of these fractions resulted in mortalities (60–100% in less than 24 h) only in fractions containing live dinospores ('whole water', 'dinoflagellate'), with no mortalities in 'cell-free' or 'bacteria'-enriched fractions. Video-micrography and electron microscopy show dinospores swarming toward and attaching to skin, actively feeding, and rapidly denuding fish of epidermis. We show here that our cultures of actively fish-killing *P. shumwayae* do not secrete potent exotoxins; rather, fish mortality results from micropredatory feeding.

Massive fish kills in mid-Atlantic USA estuaries involving several million Atlantic menhaden, *Brevoortia tyrannus*, have been attributed to dinoflagellates of the toxic *Pfiesteria* complex (TPC)⁹. Potent ichthyotoxins secreted during *Pfiesteria* blooms are thought to be responsible for mortality as well as for deeply penetrating, so-called 'Pfiesteria-specific' skin ulcers in these fish^{1,5,9}. However, earlier investigations attributed the menhaden ulcers to fungal infections^{10,11}, and *Aphanomyces invadans*, a highly pathogenic oomycete¹², is now considered the aetiological agent^{13,14}. We recently demonstrated that *A. invadans* is a primary pathogen, able to elicit menhaden ulcer disease in the absence of *Pfiesteria* species or other environmental stressors¹⁵. Thus, the role of *Pfiesteria* species in menhaden lesion events is now questioned^{13–16}.

In contrast to the oomycete-induced ulcers of wild menhaden, laboratory exposure of fishes to an unidentified *Pfiesteria* species elicited rapid, widespread epidermal erosion, osmoregulatory dysfunction and death, with potent exotoxins assumed responsible⁴. However, direct attachment of *P. shumwayae* dinospores to skin, gills, olfactory organs, the oral mucosa and the lateral line canal, associated with extensive tissue damage, has been observed¹⁶. A direct physical association with these fish tissues had not to our knowledge been previously reported, and this suggested an alternative mechanism of pathogenesis for *P. shumwayae*. To better understand this association and to clarify the consequences of dinospore attachment, we conducted laboratory challenges using a sensitive larval fish bioassay.

We exposed larval sheepshead minnows, *Cyprinodon variegatus*, to *Pfiesteria* spp. in six-well tissue culture plates containing polycarbonate membrane inserts (Millicell). This created two compartments within each well ('in', inside insert; 'out', outside insert), allowing separation of fish from dinospores across a permeable membrane (Fig. 1). Mortalities occurred only in treatments where fish and *P. shumwayae* dinospores were in direct physical contact (Fig. 2a: B in, D in, F). Fish physically separated from dinospores (A in, B out versus in) did not die, even if they resided within the same well as dying fish in contact with dinospores (B in versus out). Fish in negative controls (C) and fish exposed to a non-pathogenic strain

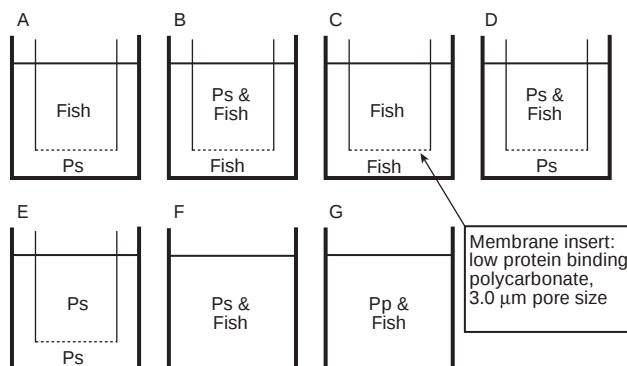


Figure 1 Experimental design for the membrane insert study using larval *Cyprinodon variegatus* exposed to *Pfiesteria shumwayae* (Ps) and *P. piscicida* (Pp).

of *P. piscicida* dinospores (G) exhibited no mortalities. When fish and *P. shumwayae* dinospores were in physical contact, mortalities began within several hours of assay initiation, reached 7.4–25.9% by 24 h and 92–100% by 48 h (Fig. 2a). Dinospore densities increased significantly after 24 h in treatments where they were in contact with fish (B in, D in, F), but remained near initial levels in treatments where they were separated from fish (A out, D out) (Fig. 2b). Reactive ammonia levels (Fig. 2c) remained below 0.25 mg l⁻¹ and dissolved oxygen concentrations (Fig. 2d) remained above 6.5 mg l⁻¹ in all treatments.

To test permeability of the polycarbonate membrane to representative lipophilic and hydrophilic algal toxins, we exposed larval fish, sequestered in clean water within membrane inserts, to a range of brevetoxin (PbTx-2) and saxitoxin (STx) concentrations placed outside the insert. Additional replicates were run without fish. After a 24-h period of exposure and equilibration, we measured toxin concentrations inside and outside the inserts. Both toxins readily diffused across the polycarbonate membrane, and rapidly equilibrated, in the presence and absence of fish (Fig. 2e, f). Fish exposed to 100 ng ml⁻¹ (nominal) PbTx-2 exhibited 100% mortality within 24 h. No mortalities occurred at lower PbTx-2 doses, in STx exposures, or in controls.

We conducted fractionation studies to investigate exotoxin secretion further and to evaluate potential contributions of con-

taminating microbial organisms and poor water quality to fish mortality. We centrifuged and filtered water taken from a 38-l bioassay of *P. shumwayae* that actively killed tilapia, *Oreochromis niloticus*. The various fractions were assayed on larval fish. Fish in 'whole water' and in the 'dinoflagellate'-enriched fractions exhibited 100% mortality within 24 and 48 h, respectively (Fig. 3a). Negligible mortality (<3.3%) occurred in the 'cell-free' (putatively 'exotoxin' enriched) and 'bacteria' fractions and controls. Dinospore densities in the whole-water and dinoflagellate fractions ranged from ~350 to ~9,000 cells ml⁻¹ during exposures (Fig. 3b). Water quality measurements were similar among all treatments. Reactive NH₃ in all treatments remained below 'high' ammonia controls (Fig. 3c). Dissolved oxygen ranged from 7.6 to 4.4 mg l⁻¹; the lowest values occurred after fish mortality in the whole-water and dinoflagellate fractions at 24 and 48 h, respectively (Fig. 3d). Similar fractionation studies with larval fish have been repeated six times with identical results (data not shown).

Microscopically, dinospores of *P. shumwayae* placed with fish exhibited rapid chemotaxis, attachment to fish epidermis by a peduncle, and feeding by myzocytosis¹⁷. Dinospores fed for about 1 min, swelling significantly as the epidermal cell cytoplasm was internalized; they then detached from fish and slowly swam away. Our *P. piscicida* cultures, which have never been pathogenic, did not exhibit attachment and feeding (see Supplementary Information).

Scanning electron microscopy (SEM) of larval fish skin after brief exposure to *P. shumwayae* showed focal epidermal degeneration associated with dinospore feeding. Myzocytosis caused contraction and sloughing of epidermal cells and cellular debris, rapidly creating shallow epidermal erosions seen only in association with numerous adhering dinospores (Fig. 4a). Dinospores attached by the distal end of the peduncle, which exhibited small filopodial extensions. The surface of affected epidermal cells was extensively damaged (Fig. 4b, c). Transmission electron microscopy (TEM) showed the distal peduncular surface of an actively feeding dinospore tightly apposed to a small group of epidermal cells, which, according to the nuclear and cytosolic morphology, were degenerating. Epidermal cell organelles, including mitochondria and rough endoplasmic reticulum, occurred within the peduncle, indicating ingestion and transport toward a large food vacuole within the dinospore epicone (Fig. 4c). Peduncles not yet attached to fish epidermis exhibited small distal

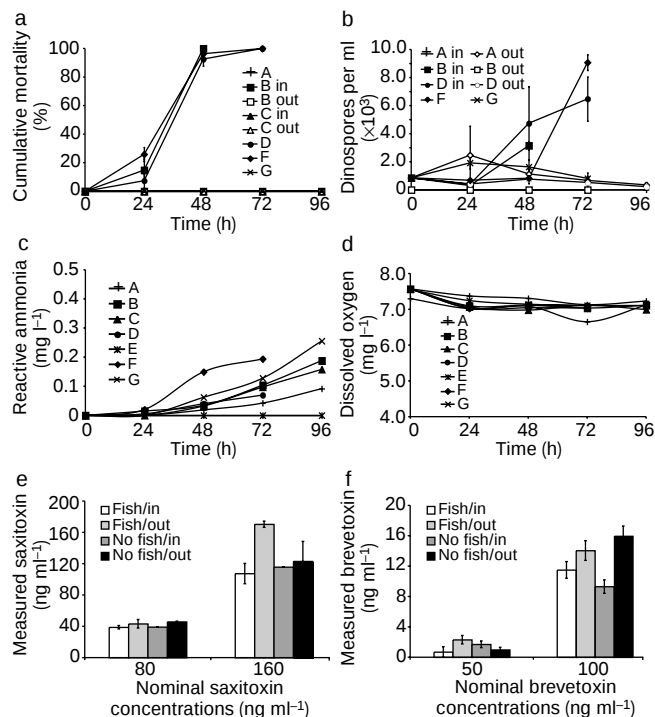


Figure 2 Results from bioassays using larval *Cyprinodon variegatus* exposed in six-well tissue culture plates fitted with polycarbonate membrane inserts. **a–d**, Mortality, cell counts and water quality from a 96-h bioassay. **e, f**, Measured saxitoxin and brevetoxin concentrations from a 24-h bioassay. **a**, Cumulative fish mortality in treatments containing fish. Survival analysis showed significant differences in mortality rates between fish in contact with dinospores and those separated from contact (Tarone-Ware $\chi^2 = 226.55$, d.f. = 7, $P < 0.001$). **b**, Cell count data from selected treatments where dinospores were in direct contact with fish and other selected treatments. **c**, Reactive ammonia concentrations for all treatments (inside and outside water samples pooled). **d**, Dissolved oxygen concentrations for all treatments (inside and outside water samples pooled). **e**, Measured saxitoxin concentrations inside and outside polycarbonate membrane inserts after 24 h, in the presence and absence of fish. **f**, Measured brevetoxin concentrations inside and outside polycarbonate membrane inserts after 24 h, in the presence and absence of fish.

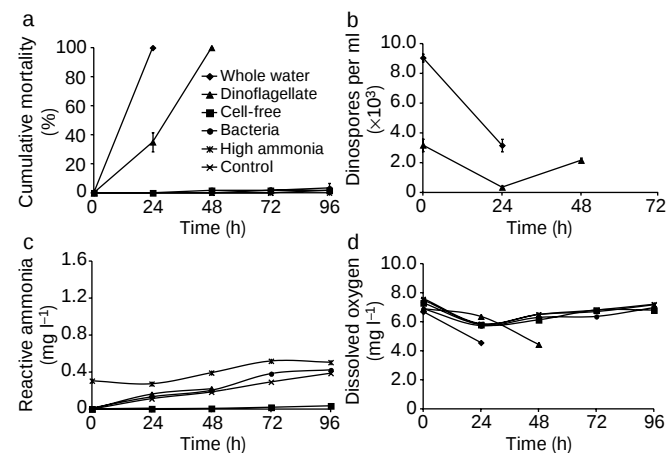


Figure 3 Fractionation assay with larval *Fundulus heteroclitus* using a *Pfiesteria shumwayae* culture that actively killed tilapia in a 38-l assay. **a**, Cumulative fish mortality during a 96-h fractionation study. Survival analysis showed significantly different mortality rates between fish exposed to fractions containing dinoflagellates and those exposed to other fractions (Tarone-Ware $\chi^2 = 302.03$, d.f. = 5, $P < 0.001$). **b**, Dinospore cell count in the 'whole water' and 'dinoflagellate' fractions. **c**, Reactive ammonia concentrations for all treatments over the 96-h exposure. **d**, Dissolved oxygen concentrations for all treatments over the 96-h exposure.

peduncular filopodia and numerous electron-dense, rod-shaped granules (Fig. 4, inset). Both structures were absent in peduncles that had attached to epidermal cells.

We report here an alternate mechanism of pathogenicity for *P. shumwayae*. In contrast to prior studies^{18,19}, our *P. shumwayae* cultures, although highly pathogenic to fish, do not produce exotoxins. Rather, they kill fish by myzocytosis, a feeding mechanism previously observed in non-toxic ambush predator dinoflagellates^{20,21} and *Pfiesteria* spp.^{22,23}, but not generally implicated in fish mortality. Despite almost a decade of research, isolation and characterization of *Pfiesteria* toxins have not been achieved. However, two distinct fractions with biological activity were recently isolated from *P. piscicida* cultures⁸. A lipophilic fraction was dominated by a phthalate ester determined to be a contaminant of artificial sea salts used to produce culture water. A hydrophilic fraction (designated pPFTx) induced a GH₄C₁ reporter gene assay, caused cell toxicity and killed brine shrimp and fish⁸. However, typical epidermal pathology attributable to *Pfiesteria* exposure¹⁶ was

not produced by these fractions. The pPFTx was also shown to interact with a P2X₇ receptor suggested to mediate GH₄C₁ cytotoxicity²⁴. Further, pPFTx selectively inhibited the *N*-methyl-D-aspartate neurotransmitter receptor, although an association with animal toxicity was not shown²⁵.

On the basis of the recently proposed nomenclature for *Pfiesteria* functional types (for example, tox-A, tox-B and non-inducible)⁵, our *P. shumwayae* cultures should be 'non-inducible' and should not kill fish. We maintained these cultures on algae for more than 24 months, presumably too long to maintain fish-killing capacity⁵. However, using standard protocols^{5,26} and environmentally relevant cell densities (300–2,500 cells ml⁻¹)²⁷, our cultures kill fish within hours (tox-A equivalent) and elicit typical pathology previously ascribed to *Pfiesteria* toxigenicity^{4,5}. Thus, our findings challenge the utility of what has been described as the 'gold standard' 38-l *Pfiesteria* toxicity bioassay^{19,26}, which uses fish neurobehavioural changes, gross pathology and mortality as sole endpoints of 'toxigenicity'. That assay cannot discriminate between truly toxigenic cultures and cultures that kill fish by myzocytosis. Our findings indicate that fish mortality after *P. shumwayae* exposure results from micropredatory feeding on captive fish prey. Certainly, myzocytosis resulting in widespread skin damage and rapid mortality can convincingly be demonstrated *in vitro*. However, what role, if any, *Pfiesteria* spp. have in morbidity and mortality of wild fishes has recently been questioned^{13–16,28,29} and remains unclear. Only assays that routinely apply fractionation or membrane isolation protocols will discriminate between disparate killing mechanisms. We therefore recommend that all strains and cultures of *Pfiesteria* be re-evaluated for pathogenicity by these protocols. □

Methods

Pfiesteria culture

The clonal culture of *Pfiesteria shumwayae* (VIMS-1049) was established from a water sample obtained from the Pamlico River, North Carolina, on 12 November 1999. It has been deposited with the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (CCMP number 2089). Dinoflagellate species were definitively identified by SEM plate tabulations^{18,30} and by comparing DNA sequence analysis of the small subunit ribosomal RNA gene with GenBank deposited sequences (*P. piscicida* accession number AF149793, *P. shumwayae* AF080098). Culture identities were intermittently re-verified using species-specific polymerase chain reaction primers described in Supplementary Information.

Membrane insert study

The experimental design illustrated in Fig. 1 prevented dinospores from making physical contact with fish, yet allowed exotoxins, if present, to diffuse across a membrane and affect fish. We used membrane inserts that fit into wells of tissue culture plates (see below). Cultures of *P. shumwayae* (VIMS-1049) and *P. piscicida* (VIMS-P11) were used at a density of ~1,000 cells ml⁻¹ in sterile-filtered autoclaved (12%) York River water supplemented with penicillin (2,500 IU l⁻¹) and streptomycin (2.5 mg l⁻¹).

Membrane permeability study

Brevetoxin extracts used in the exposures were produced from a *Karenia brevis* strain isolated from the 1999 bloom at Pensacola Beach, Florida, and maintained at the Environmental Protection Agency (EPA) Laboratory, Gulf Breeze, Florida. Brevetoxin isolation protocols are described in Supplementary Information. Brevetoxin extract (1 vial: 50 µg PbTx-2) was solubilized in 100% methanol (0.5 ml), dissolved into 500 ml of 12% artificial sea water (ASW) and serially diluted to obtain the range of concentrations required. A methanol control (0.1% v/v) was included.

Saxitoxin (Sigma; 10 µg) was solubilized in 0.5 ml of ASW (12%) and then dissolved in 62 ml (62.5 ml total volume) of ASW to obtain a high-dose stock solution (160 ng ml⁻¹). We distributed 30 ml of this stock to six wells (5 ml well⁻¹), and serially diluted 30 ml to obtain the lower concentrations.

The assay was conducted in six-well tissue culture plates (Falcon) over a range of concentrations (PbTx-2, 0, 10, 25, 50, 100 ng ml⁻¹; STx, 0, 20, 40, 80, 160 ng ml⁻¹). A membrane insert (Millicell, polycarbonate, pore size 3 µm) was set into each well and 5 ml of clean filtered ASW (12%) was immediately added to the insert. Three replicates per dose received five larval fish within the insert, whereas three replicates per dose remained without fish. The assay was run for 24 h in a BSL2 cabinet at room temperature (22 °C) with mortalities recorded at the end of the study. Brevetoxin and saxitoxin analyses are detailed in Supplementary Information.

Fractionation study

To identify and select cultures, we obtained materials for fractionation studies from 'standard' 38-l bioassay aquaria¹⁹ that exhibited high mortalities of tilapia (20–30% daily).

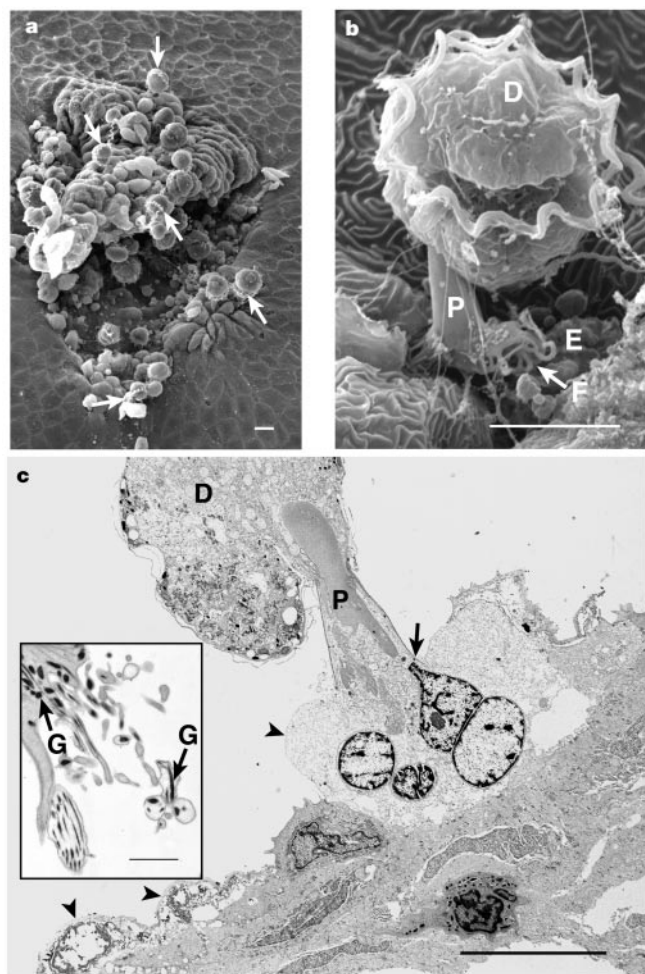


Figure 4 Myzocytosis by *Pfiesteria shumwayae* on the epidermis of larval *Cyprinodon variegatus*. **a**, Scanning electron micrograph (SEM) of early epidermal erosion (<5 min) with numerous attached dinospores (arrows). **b**, SEM of single dinospore (D) attached to and feeding on damaged epidermal cell (E). Note the distal end of peduncle (P) exhibiting filopodial projections (F). **c**, Transmission electron micrograph (TEM) of dinospore (D) attachment (arrow) and feeding on epidermal cells. Note the epidermal cytoplasm and organelles within the dinospore peduncle (P), and the cytoplasmic rarefaction and nuclear degeneration of affected cells (arrowheads). Inset, TEM of distal portion of an unattached peduncle showing filopodial extensions and rod-shaped electron-dense granules (G). Scale bars: **a**, **c**, 10 µm; **b**, 5 µm; inset, 1 µm. Two short video clips of myzocytosis by *P. shumwayae* on larval fish skin (~1 min) can be seen in Supplementary Information.

Protocols for the 38-l bioassays are detailed in Supplementary Information. Water was taken from an aquarium previously amended with an environmental water sample from Slocum Creek, North Carolina, and verified by cell counts, SEM and molecular analyses to contain high densities of *P. shumwayae*.

Water from the actively killing tank was used as a positive ('whole water') control. An enriched 'dinoflagellate' fraction was produced by filtering, rinsing and resuspending 3 l of whole water. Viability of dinoflagellates in this fraction was confirmed microscopically. A 'bacteria' fraction was obtained by centrifuging 3 l of whole water at 8,500 g for 45 min at 10 °C, resuspending the pellet in 12‰ ASW, filtering through a 5-µm filter to remove dinoflagellates and other protozoa, and bringing the filtrate to 3 l using 12‰ ASW. A 'cell-free' fraction was obtained by removing the supernatant from the centrifuged bacterial pellet and filtering through a 5-µm and then a 0.45-µm filter to a volume of 3 l. Filter-sterilized 12‰ ASW was used as a negative control. We also used a 'high ammonia' control consisting of 12‰ ASW with ammonia and pH adjusted to that of the whole water.

Electron microscopy

Fish killed with trichaine methanesulphonate (MS-222) were fixed in 4% glutaraldehyde with 5% paraformaldehyde in 0.1 M sodium cacodylate buffer at pH 7.2, at room temperature for ~2 h. Samples were washed with three changes of 0.1 M sodium cacodylate buffer, 15–30 min each, and stored overnight at 4 °C in a third change of buffer. They were postfixed with 1% OsO₄ in 0.1 M buffer, pH 7.2, at room temperature for 1 h and then washed with buffer in three changes of 0.1 M cacodylate, pH 7.2, 15–30 min each. The caudal peduncle was cut off the fish with a single-edged razor blade and processed for TEM analysis, and the body of each fish was processed for SEM by standard methods detailed in Supplementary Information.

Other methods

Larval fish sources, water quality measurements and cell count protocols are detailed in Supplementary Information.

Received 20 June; accepted 24 July 2002; doi:10.1038/nature01008.

Published online 5 August 2002.

- Burkholder, J. M., Noga, E. J., Hobbs, C. H. & Glasgow, H. B. Jr New "phantom" dinoflagellate is the causative agent of major estuarine fish kills. *Nature* **358**, 407–410 (1992).
- Burkholder, J. M., Glasgow, H. B. Jr & Hobbs, C. B. Fish kills linked to a toxic ambush-predator dinoflagellate: distribution and environmental conditions. *Mar. Ecol. Prog. Ser.* **124**, 43–61 (1995).
- Glasgow, H. B. Jr, Burkholder, J. M., Schmechel, D. E., Tester, P. E. & Rublee, P. A. Insidious effects of a toxic estuarine dinoflagellate on fish survival and human health. *J. Toxicol. Environ. Health* **46**, 501–522 (1995).
- Noga, E. J., Khoo, L., Stevens, J. B., Fan, Z. & Burkholder, J. M. Novel toxic dinoflagellate causes epidemic disease in estuarine fish. *Mar. Pollut. Bull.* **32**, 219–224 (1996).
- Burkholder, J. M., Glasgow, H. B. Jr & Deamer-Melia, N. J. Overview and present status of the toxic *Pfiesteria* complex (Dinophyceae). *Phycologia* **40**, 186–214 (2001).
- Grattan, L. M. *et al.* Learning and memory difficulties after environmental exposure to waterways containing toxin-producing *Pfiesteria* or *Pfiesteria*-like dinoflagellates. *Lancet* **352**, 532–539 (1998).
- Grattan, L. M., Oldach, D. & Morris, G. Human health risks of exposure to *Pfiesteria piscicida*. *Bioscience* **51**, 853–857 (2001).
- Moeller, P. D. *et al.* Current progress in isolation and characterization of toxins isolated from *Pfiesteria piscicida*. *Environ. Health Perspect.* **109**, 739–743 (2001).
- Glasgow, H. B., Burkholder, J. M., Mallin, M. A., Deamer-Melia, N. J. & Reed, R. E. Field ecology of toxic *Pfiesteria* complex species and a conservative analysis of their role in estuarine fish kills. *Environ. Health Perspect.* **109**, 715–730 (2001).
- Noga, E. J., Levine, J. F., Dykstra, M. J. & Hawkins, J. H. Pathology of ulcerative mycosis in Atlantic menhaden. *Dis. Aquat. Org.* **4**, 189–197 (1988).
- Dykstra, M. J. *et al.* Ulcerative mycosis: a serious menhaden disease of the southeastern coastal fisheries of the United States. *J. Fish Dis.* **12**, 175–178 (1989).
- Callinan, R. B. in *Proc. ODA Regional Sem. Epizootic Ulcerative Syndrome* (eds Roberts, R. J., Campbell, B. & MacRae, I. H.) 248–252 (The Aquatic Animal Health Research Institute, Bangkok, 1994).
- Blazer, V. S. *et al.* *Aphanomyces* as a cause of ulcerative skin lesions of menhaden from Chesapeake Bay tributaries. *J. Aquat. Anim. Health* **11**, 340–349 (1999).
- Blazer, V. S. *et al.* *Aphanomyces invadans* in Atlantic menhaden along the East Coast of the United States. *J. Aquat. Anim. Health* **14**, 1–10 (2002).
- Kiryu, Y., Shields, J. D., Vogelbein, W. K., Zwerner, D. E. & Kator, H. Induction of skin ulcers in Atlantic menhaden by injection and water-borne exposure to the zoospores of *Aphanomyces invadans*. *J. Aquat. Anim. Health* **14**, 11–24 (2002).
- Vogelbein, W. K., Shields, J. D., Haas, L. W., Reece, K. S. & Zwerner, D. E. Skin ulcers in estuarine fishes: A comparative pathological evaluation of wild and laboratory-exposed fish. *Environ. Health Perspect.* **109**, 687–693 (2001).
- Schnepf, E. & Deichgräber, G. "Myxocytosis", a kind of endocytosis with implications on compartmentation in endosymbiosis. Observations in *Paulsenella* (Dinophyta). *Naturwissenschaften* **71**, 218–219 (1983).
- Glasgow, H. B., Burkholder, J. M., Morton, S. L. & Springer, J. A second species of ichthyotoxic *Pfiesteria* (Dinamoebales, Dinophyceae). *Phycologia* **40**, 234–245 (2001).
- Burkholder, J. M., Marshall, H. G., Glasgow, H. B., Seaborn, D. W. & Deamer-Melia, J. The standardized fish bioassay procedure for detecting and culturing actively toxic *Pfiesteria*, used by two reference laboratories for Atlantic and Gulf Coast states. *Environ. Health Perspect.* **109**, 745–756 (2001).
- Spero, H. J. & Moree, M. D. Phagotrophic feeding and its importance to the life cycle of the holozoic dinoflagellate, *Gymnodinium funigiforme*. *J. Phycol.* **17**, 43–51 (1981).
- Schnepf, E. & Elbrächter, M. Nutritional strategies in dinoflagellates—a review with emphasis on cell biological aspects. *Eur. J. Protistol.* **28**, 3–24 (1992).
- Vogelbein, W. K. *et al.* in *Agenda: Symp. Harmful Marine Algae in the U.S.* 81 (The Marine Biological Laboratory, Woods Hole, Massachusetts, 2000).

- Shumway, S., Springer, J., Burkholder, J. M., Glasgow, H. B. in *Agenda: Symp. Harmful Marine Algae in the U.S.* 64 (The Marine Biological Laboratory, Woods Hole, Massachusetts, 2000).
- Melo, A. C., Moeller, P. R., Glasgow, H. B., Burkholder, J. M. & Ramsdell, J. S. Microfluorimetric analysis of a purinergic receptor (P2X₇) in GH₄C₁ rat pituitary cells: Effects of a bioactive substance produced by *Pfiesteria piscicida*. *Environ. Health Perspect.* **109**, 731–737 (2001).
- El-Nabawi, A. *et al.* The N-methyl-D-aspartate neurotransmitter receptor is a mammalian brain target for the dinoflagellate *Pfiesteria piscicida* toxin. *Toxicol. Appl. Pharmacol.* **169**, 84–93 (2000).
- Marshall, H. M. *et al.* Comparative culture and toxicity studies between the toxic dinoflagellate, *Pfiesteria piscicida*, and a morphologically similar cryptoperidiniopsis dinoflagellate. *J. Exp. Mar. Biol. Ecol.* **255**, 65–74 (2000).
- Burkholder, J. M. & Glasgow, H. B. *Pfiesteria piscicida* and other toxic *Pfiesteria*-like dinoflagellates: behaviour, impacts and environmental controls. *Limnol. Oceanogr.* **42**, 1052–1075 (1997).
- Litaker, R. W., Vandersea, M. W., Kibler, S. R., Madden, V. J., Noga, E. J. & Tester, P. A. Life cycle of the heterotrophic dinoflagellate *Pfiesteria piscicida* (Dinophyceae). *J. Phycol.* **38**, 442–463 (2002).
- Berry, J. P. *et al.* Are *Pfiesteria* species toxicogenic? Evidence against production of ichthyotoxins by *Pfiesteria shumwayae*. *Proc. Natl Acad. Sci. USA* (in the press).
- Steidinger, K. A. *et al.* *Pfiesteria piscicida* gen. et. sp. nov. (Pfiesteriaceae fam. nov.), a new toxic dinoflagellate with a complex life cycle and behaviour. *J. Phycol.* **32**, 157–164 (1996).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

C. Squyers, A. Miller, P. Blake, D. Zwerner, E. Westcott, V. Foster, L. Ott, E. Walker, J. Watts, W. Ribeiro and K. Hudson contributed significantly to this investigation. E. M. Burreson provided support and critical review. This work was funded in part by ECOHAB (Ecology and Oceanography of Harmful Algal Blooms) grants from the EPA, the EPA with the National Oceanic and Atmospheric Administration (NOAA), and from the NOAA; the Commonwealth of Virginia's *Pfiesteria* Initiative; and the Centers for Disease Control and Prevention. We thank T. Shedd and the US Army Center for Environmental Health Research for the lease of the BSL3 laboratory.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.K.V. (e-mail: wolf@vims.edu).

Protein phosphatase 1 is a molecular constraint on learning and memory

David Genoux*, Ursula Haditsch*, Marlen Knobloch*, Aubin Michalon*, Daniel Storm† & Isabelle M. Mansuy*

* Institute of Cell Biology, Swiss Federal Institute of Technology, Department of Biology, ETH Hönggerberg, CH-8093 Zürich, Switzerland

† Department of Pharmacology, PO Box 357280, University of Washington, Seattle, Washington 98195, USA

Repetition in learning is a prerequisite for the formation of accurate and long-lasting memory. Practice is most effective when widely distributed over time, rather than when closely spaced or massed. But even after efficient learning, most memories dissipate with time unless frequently used^{1,2}. The molecular mechanisms of these time-dependent constraints on learning and memory are unknown. Here we show that protein phosphatase 1 (PP1) determines the efficacy of learning and memory by limiting acquisition and favouring memory decline. When PP1 is genetically inhibited during learning, short intervals between training episodes are sufficient for optimal performance. The enhanced learning correlates with increased phosphorylation of cyclic AMP-dependent response element binding (CREB) protein, of Ca²⁺/calmodulin-dependent protein kinase II (CaMKII) and of the GluR1 subunit of the AMPA receptor; it also correlates with CREB-dependent gene expression that, in control mice, occurs only with widely distributed training. Inhibition of PP1 prolongs memory when induced after learning, suggesting that PP1 also promotes forgetting. This property may account for

ageing-related cognitive decay, as old mutant animals had preserved memory. Our findings emphasize the physiological importance of PP1 as a suppressor of learning and memory, and as a potential mediator of cognitive decline during ageing.

One feature of many forms of vertebrate learning is that practice determines whether learned material is remembered or forgotten, and whether the retained information is precise or vague. During practice, the pattern of distribution of training sessions is critical, and long intervals between sessions allow better encoding and more robust memory than do short intervals¹. A possible explanation for this interval-dependent phenomenon is that with frequent and repetitive training, the processing of information from a trial is attenuated because information from the previous trial is still being processed, suggesting that a minimal delay is required for complete input integration². An additional temporal limit on learning and memory is that much of the information that is stored after learning gradually dissipates with time unless regularly retrieved and re-consolidated³. These time-dependent constraints on learning and memory are poorly understood.

Several reports have suggested the existence of endogenous

molecular suppressors that negatively control the efficacy of neuronal transmission and memory formation⁴. One proposed category of such memory suppressors is constituted by protein phosphatases. These molecules, together with protein kinases, regulate many cellular processes by the reversible phosphorylation/dephosphorylation of specific substrates⁵. For instance, the Ca^{2+} /calmodulin-dependent protein phosphatase calcineurin (PP2B) was recently reported to block learning, memory storage and memory retrieval⁶⁻⁸. Here we examine the role of the protein phosphatase 1 (PP1), a phosphatase thought to act downstream from calcineurin and to negatively regulate synaptic plasticity^{9,10} in the mechanisms that constrain learning efficacy and that underlie forgetting.

To determine whether delays between training sessions influence learning in rodents, we tested mice in an object recognition task¹¹. In this task, memory traces of different strength can be elicited by varying the number and duration of training sessions. Memory for objects is inferred from the animal's ability to distinguish a novel

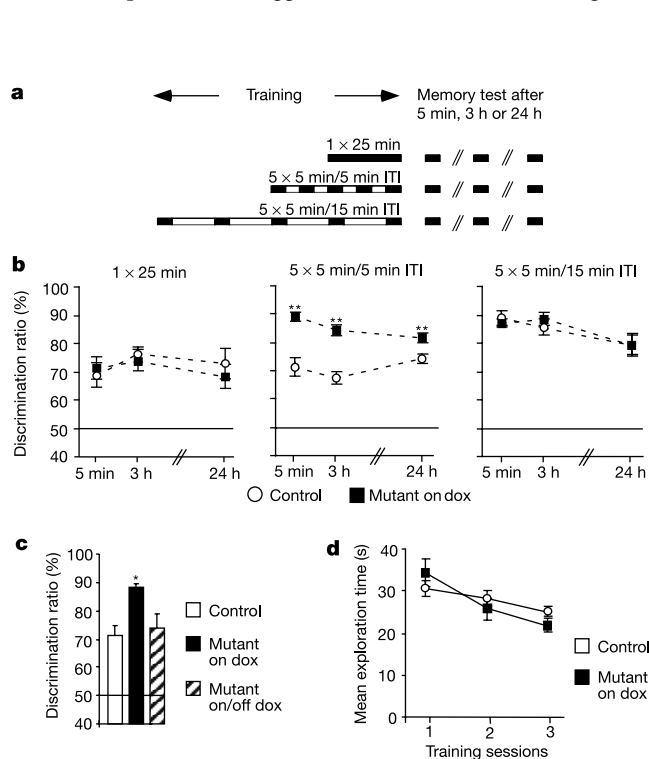


Figure 1 Distributed training improves performance in the object recognition task. **a**, Scheme of training and testing protocols. ITI, inter-trial interval. **b**, Discrimination ratios in control mice 5 min ($F(1,23) = 10.296$, $P < 0.01$, for 5-min versus 15-min ITI), 3 h ($F(1,9) = 22.976$, $P = 0.001$ for 5-min versus 15-min ITI) or 24 h after training with either one of the protocols described in **a**. Discrimination ratios in mutant mice 5 min ($F(1,13) = 16.565$, $P = 0.001$ for 1 x 25-min versus 5 x 5 min/5-min ITI), 3 h ($F(1,13) = 5.217$, $P < 0.05$ for 1 x 25-min versus 5 x 5 min/5-min ITI), or 24 h ($F(1,20) = 14.823$, $P = 0.01$ for 1 x 25-min versus 5 x 5 min/5-min ITI) after training. Lines at 50 indicate chance level. Dox, doxycycline. **c**, Reversibility of the enhancement in performance. Mutant mice ($n = 9$) were treated with dox, trained with a set of objects for five 5-min sessions with 5-min ITIs and tested 5 min later (Mutant on dox). Dox was withdrawn and the animals were trained 2 weeks later with five 5-min sessions/5-min ITIs using a second set of objects, then tested 5 min later (Mutant on/off dox; $F(1,14) = 6.816$, $P < 0.05$ mutant on dox versus on/off dox). Control mice $n = 22$. **d**, Representative graph of total exploration during training in control ($n = 22$) and mutant ($n = 7$) mice. Decrease in total exploration per object across training sessions (1 to 3) indicates habituation to the objects. Control groups in these experiments and the following ones were treated or not treated with dox. Results from these groups were similar and therefore were pooled.

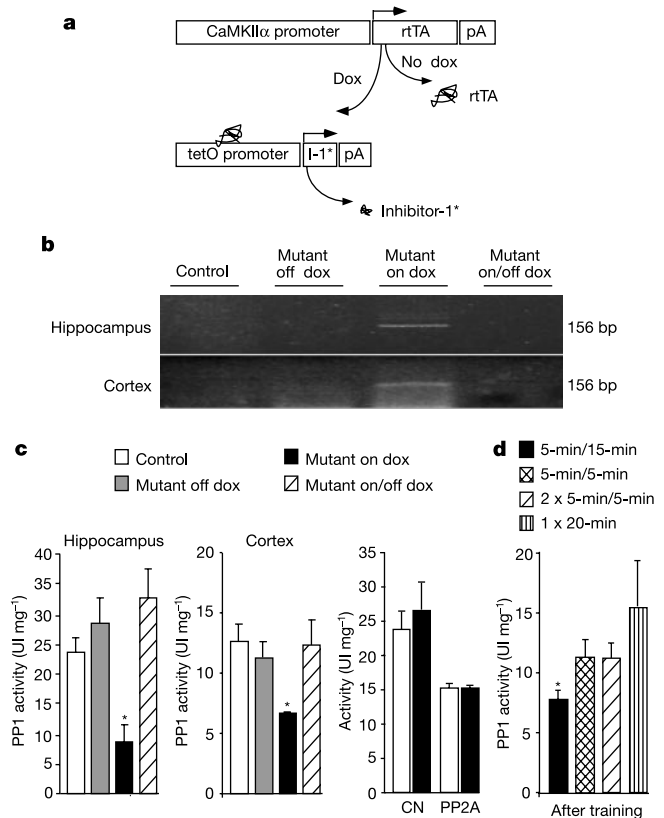


Figure 2 Genetic or training-dependent inhibition of PP1. **a**, Schematic representation of the transgenes used to achieve dox-dependent *I-1** expression. Heterozygous mice expressing rtTA with the CaMKII α promoter were crossed with heterozygous mice carrying the *tetO* promoter fused to the *I-1** gene. In double transgenic animals, dox induces *I-1** expression while no expression occurs in the absence of dox. pA, polyadenylation signal. **b**, Dox-dependent and reversible *I-1** mRNA expression. A 156-base-pair (156 bp) *I-1** band was detected by RT-PCR in hippocampus and cortex from mutant mice on dox. There was no band in the absence of dox (Mutant off dox) or two weeks after dox removal (Mutant on/off dox). **c**, Dox-dependent and reversible inhibition of PP1. PP1 activity is reduced in hippocampus and cortex of mutant mice on dox ($n = 3$) compared to control mice ($n = 6$), mutant mice off dox ($n = 4$) or mutant mice on/off dox ($n = 3$) ($P < 0.05$ for all). Right panel, calcineurin (CN) and PP2A activity in control and mutant mice. UI, nmol phosphate released per min per mg protein. **d**, Training-dependent inhibition of PP1. PP1 activity is reduced in cortex of control mice trained for 5 min followed by a 15-min delay ($n = 6$, versus naive or all other trained groups, $P < 0.05$) but not in mice trained once for 5 min followed by a 5-min delay ($n = 4$) or twice for 5-min followed by a 5-min delay (20 min in total, $n = 3$), or trained once for 20 min ($n = 3$).

object from familiar objects after learning. Mice were trained on a massed (a single 25-min trial) or a distributed protocol with brief intervals (five 5-min trials with 5-min inter-trial intervals, ITIs) then tested for their memory for objects (Fig. 1a). In control mice, memory performance was moderate when tested 5 min ($n = 7$), 3 h ($n = 9$) or 24 h ($n = 8$) after massed training (open circles in left panel, Fig. 1b). Similarly, distributed training with brief ITIs elicited

average discrimination ratios after 5 min ($n = 10$), 3 h ($n = 6$) or 24 h ($n = 13$) ($P > 0.1$ versus massed training, middle panel, Fig. 1b), denoting the limited efficacy of massed or brief-interval training. To test whether longer delays between training trials would be more efficient, ITIs were extended from 5 to 15 min. These longer intervals led to enhanced performance whether the animals were tested 5 min ($n = 7$), 3 h ($n = 6$) or 24 h ($n = 6$) after training ($P < 0.01$ versus 5-min ITIs, right panel, Fig. 1b).

To assess if this effect may result from the relief of a memory suppressor and examine whether PP1 could be such a suppressor, PP1 activity was genetically reduced in the brain of transgenic mice and the effect on performance was evaluated. Mutant mice expressing inducibly a constitutively active form of inhibitor 1 ($I-1^*$)¹², an endogenous inhibitor of PP1 (ref. 13), were generated using the reversible tetracycline-controlled transactivator (rtTA) system under the control of the Ca^{2+} /calmodulin-dependent protein kinase α (CaMKII α) promoter^{7,14} (Fig. 2a). In double transgenic animals, $I-1^*$ messenger RNA was expressed in hippocampus and cortex (Fig. 2b) and PP1 activity was reduced by respectively $67.7 \pm 12\%$ and $45.5 \pm 1\%$ in these brain areas upon doxycycline (dox) treatment ($P < 0.05$, Fig. 2c). No mRNA expression (Fig. 2b) and no reduction in PP1 activity (Fig. 2c) were detected in mutant animals not treated with dox. Moreover, transgene mRNA expression (Fig. 2b) and PP1 inhibition (Fig. 2c) were fully reversible and could be suppressed in both adult hippocampus and cortex after dox removal. Phosphatase inhibition was specific to PP1 as calcineurin or PP2A activity was not changed by transgene expression in the mutant mice (Fig. 2c).

In $I-1^*$ mutant mice, massed training in the object recognition task led to moderate performance 5 min ($n = 8$), 3 h ($n = 8$) or 24 h ($n = 8$) after training, similarly to control mice (filled squares in left panel, Fig. 1b). By contrast with closely spaced sessions, $I-1^*$ mutant mice had significantly enhanced memory for objects at all intervals after training (5 min and 3 h, $n = 7$; 24 h, $n = 14$, $P < 0.01$ versus control, middle panel, Fig. 1b). Indeed, performance was similar to that achieved in control mice trained with widely spaced sessions and was optimal, as widely distributed training in $I-1^*$ mutant mice did not further increase discrimination (5 min, $n = 9$; 3 h, $n = 8$; 24 h, $n = 8$; right panel, Fig. 1b). This improvement was directly due to PP1 inhibition and not to any permanent alteration resulting from $I-1^*$ transgene expression, as it was fully reversed by suppressing transgene expression and retraining the animals with different objects (Mutant on/off dox, Fig. 1c). Moreover, this reversion was not due to prior learning as two consecutive sessions of training led to similar performance on both sessions in control mice. Furthermore, in mutant mice, transgene expression induced only after the second session (mutant off/on) enhanced performance to a similar extent as when induced before the first session (not shown). Finally, total exploration of objects during training was similar in all groups, reflecting comparable motivation (Fig. 1d). These results therefore indicate that both the genetic inhibition of PP1 and long ITIs in training favoured learning and memory in a similar and mutually occlusive manner. To confirm this effect, we examined whether long ITIs in control mice correlated with PP1 inhibition. PP1 activity was consistently and significantly reduced in control animals trained for 5 min followed by a 15-min delay, but not in animals trained for one or two sessions of 5 min followed by a 5-min delay or trained on a massed 20-min session (Fig. 2d), supporting the hypothesis that widely distributed training is more efficient because it relieves a PP1-dependent constraint. This profound effect of training on PP1 may reflect the need for rapid neutralization of PP1, possibly together with activation of opposing kinases, for optimal signal transduction.

To identify the signalling pathways involved and determine whether PP1 inhibition recruits similar mechanisms whether induced genetically or with long ITIs, we examined several targets of PP1. As the formation of long-term memory requires protein

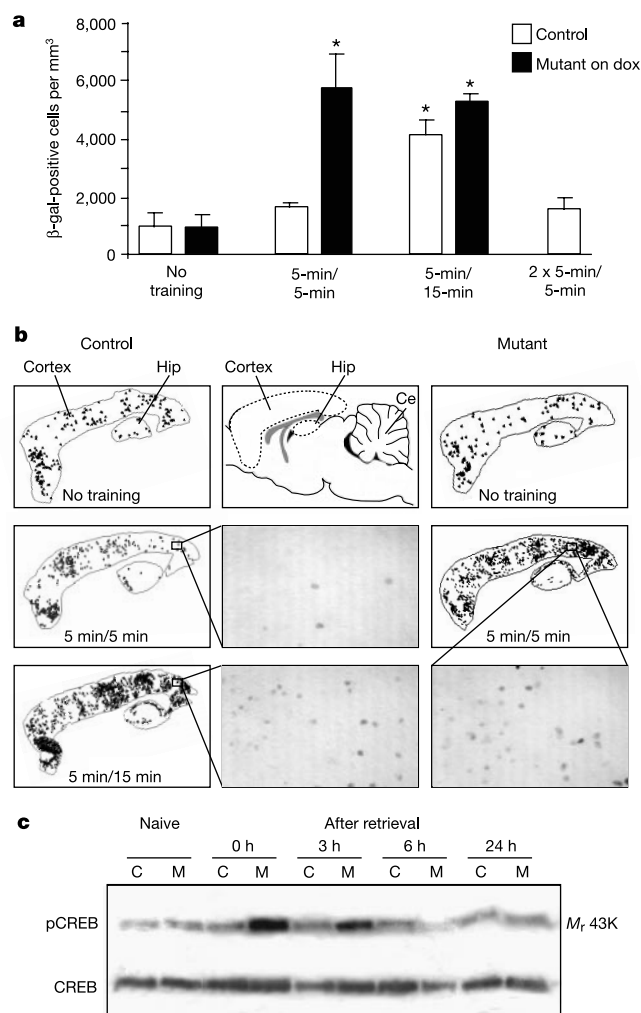


Figure 3 Enhanced CREB activity and phosphorylation. **a**, CREB transcriptional activity during training. Bar graphs are mean number of β -galactosidase positive cells per volume of cortex before training (no training, control, $n = 4$; mutant on dox, $n = 3$), and either after 5-min training followed by a 5-min delay (5-min/5-min, control, $n = 3$; mutant on dox, $n = 3$), 5-min training followed by a 15-min delay (5-min/15-min, control, $n = 3$; mutant on dox, $n = 3$), or two sessions of 5-min training followed by a 5-min delay (2 x 5-min/5-min, control, $n = 3$). **b**, Representative diagrams and corresponding close-ups of cell counts in control (left panels) and mutant (right panels) mice before training (no training), or after 5-min training followed by a 5-min delay or 5-min training followed by a 15-min delay (only control). Middle top panel, scheme of a mouse brain with cortex and hippocampus (Hip) lined up such as displayed in left and right diagrams. Ce, cerebellum. **c**, Time course of CREB phosphorylation after retrieval. Representative western blot of pCREB in control (C) and mutant mice on dox (M) not trained (naive) or trained on five sessions of 5-min with 5-min ITIs and tested for retrieval 24 h later. The levels of pCREB and total CREB (relative molecular mass $M_r = 43,000$) were examined either immediately, 3, 6 or 24 h after retrieval (both control and mutant mice, $n = 2$ for each time point). In the mutant mice, pCREB level was high immediately after, and 3 h after, memory retrieval, then dropped back to baseline after 6 h. The pCREB level was only slightly increased in control mice after 3 and 6 h. The level of total CREB protein was similar in control and mutant animals.

synthesis and involves primarily the transcription factor CREB protein^{15,16}, a PP1 substrate¹⁷, we first examined CREB activity during learning. Hence, we crossed the *I-1** animals with reporter mice carrying CREB-responsive elements (Cre) fused to the *lacZ* gene in which CREB-mediated gene expression can be monitored by β -galactosidase staining¹⁸. Before training, β -galactosidase expression was low in *I-1** control and mutant mice carrying the *lacZ* gene (Fig. 3a, b), indicating minimal CREB activation in resting conditions. By contrast, in mutant but not in control mice, 5-min training followed by a 5-min delay strongly increased CREB-dependent gene expression in both cortex ($F(1, 4) = 13.41$, $P < 0.05$, Fig. 3a, b) and hippocampus (bar graph not shown but see Fig. 3b). Likewise, a 15-min delay following training induced an intense signal in mutant mice ($F(1, 4) = 153.62$, $P < 0.01$). A similarly strong activation was achieved in control mice only when training was followed by a 15-min delay ($F(1, 4) = 28.828$, $P < 0.01$ versus 5-min delay, Fig. 3a, b). This activation was directly dependent on training distribution rather than on the sole passage of time after training, as two consecutive sessions of 5-min training/5-min delay (20 min total) did not induce more gene expression than a single such session in control mice (Fig. 3a). Finally, memory retrieval also appeared to be constrained by PP1, because the levels of phosphorylated CREB (pCREB), a marker for CREB activation¹⁹, were significantly increased in the mutant mice after retrieval

(Fig. 3c). Overall, these results indicate that both the genetic inhibition of PP1 and long delays during training increase CREB transcriptional activity in the brain, suggesting that PP1 and short intervals constrain learning by preventing gene expression. These data are consistent with previous studies in *Drosophila* and mice showing that CREB is important for determining the number of trials and the trial intervals necessary for the formation of long-term memory¹⁶.

Once learned and stored, information must be maintained in memory over time for later recall. To assess whether PP1 can also influence the persistence of memory, we tested spatial memory in the *I-1** mice using a water maze²⁰. In this maze, animals have to find an escape platform hidden in a tank of opaque water by using spatial cues placed in the experimental room. Trial after trial, new spatial information must be learned, remembered and retrieved to accurately navigate to the platform. Memory for the platform was tested at various delays after training. Consistent with the improvement in the object recognition test, spatial performance in the mutant mice was enhanced on the water maze. During training, the mutant mice located the platform faster and reached minimal escape latencies sooner than control mice (five versus seven training blocks, $F(1,274) = 10.96$ $P < 0.01$ overall, Fig. 4a), indicating that the *I-1** transgene enhanced spatial learning and memory.

In an attempt to relate this effect to a biochemical modification,

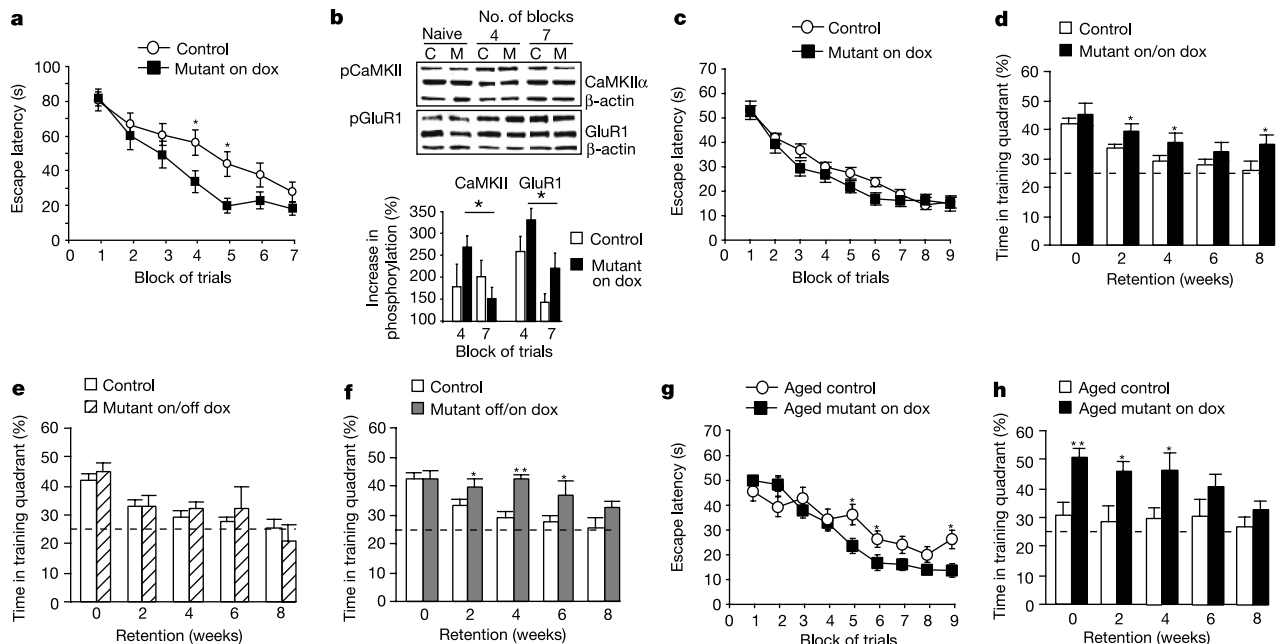


Figure 4 Improved spatial learning and memory in *I-1** mutant mice. **a**, Spatial learning in control mice ($n = 10$) and mutant mice on dox ($n = 10$) trained on seven consecutive blocks of two trials (90 seconds each) separated by 30–60 min. **b**, CaMKII and GluR1 phosphorylation in spatial training. Control and mutant mice on dox were trained as in **a**, and the levels of pCaMKII and pGluR1 were analysed by western blotting of membrane-enriched hippocampal preparations either before training (naive, CaMKII: control (C), $n = 2$; mutant (M), $n = 2$; GluR1: C, $n = 4$; M, $n = 3$) or 10 min after 4 training blocks (CaMKII: C, $n = 2$; M, $n = 3$; GluR1: C, $n = 4$; M, $n = 3$) or 7 training blocks (CaMKII: C, $n = 3$; M, $n = 2$; GluR1: C, $n = 3$; M, $n = 3$) (top panels). Total CaMKII α and GluR1, and β -actin were measured for quantification. Per cent increase in CaMKII and GluR1 phosphorylation in naive mice versus trained mice after 4 and 7 training blocks (bar graphs). **c**, Spatial acquisition in control mice ($n = 28$) and mutant mice on dox ($n = 27$) trained on nine consecutive blocks of three trials (60 seconds each) separated by 24 h. This training phase was followed by a 7-day training period using another platform position (not shown). **d–f**, Spatial memory in control and mutant mice treated with dox either **d**, during and after training (on/on dox); **e**, only during training (on/off); or **f**, only

after training (off/on) and probe trials were performed 3 h (time point 0, control, $n = 28$; mutant on dox, $n = 28$; mutant off/on dox, $n = 13$; mutant on/off dox, $n = 6$), 2 weeks (control, $n = 28$; mutant on dox, $n = 19$; mutant off/on dox, $n = 13$; mutant on/off dox, $n = 6$), 4 weeks (control, $n = 28$; mutant on dox, $n = 16$; mutant off/on dox, $n = 11$; mutant on/off dox, $n = 6$), 6 weeks (control, $n = 28$; mutant on dox, $n = 15$; mutant off/on dox, $n = 13$; mutant on/off dox, $n = 6$), and 8 weeks (control, $n = 24$; mutant on dox, $n = 14$; mutant off/on dox, $n = 13$; mutant on/off dox, $n = 5$) after training. In mutant on/off dox, behavioural testing was performed two weeks after dox removal. Dashed lines indicate chance level (25%). **g**, **h**, Spatial performance in aged mice. Learning curve (**g**), and probe trials (**h**), in aged control ($n = 9$) and mutant ($n = 9$) mice trained on nine consecutive blocks of three trials separated by 24 h (no platform transfer) and treated with dox during and after training. Probe trials were performed 1 day (time point 0), 2 weeks, 4 weeks, 6 weeks and 8 weeks after training. For all experiments, the time spent in other quadrants of the maze was similar in all groups and not significantly different from chance (not shown). Asterisk, $P < 0.05$; double asterisk, $P < 0.01$.

we examined the levels of phosphorylation of another two targets of PP1, CaMKII and the GluR1 subunit of the AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptor^{21–24}. Both are present in synapses and therefore are more likely to be activated during short training trials (like the water maze trials) than are late components of the transcriptional machinery in the nucleus. Western blot analyses revealed that the levels of both CaMKII phosphorylated at Thr 286 (pThr286CaMKII, referred to as pCaMKII) and GluR1 phosphorylated at Ser 845 (pSer845GluR1, referred to as pGluR1) increased after training in control and mutant mice ($P < 0.05$, Fig. 4b). In the mutant mice, the increase in both pCaMKII and pGluR1 was larger than in control mice after training (genotype effect, $P \leq 0.05$). These results suggest a correlation between CaMKII and GluR1 phosphorylation and performance, and a possible involvement of these PP1 targets in the observed improvement.

In order to examine long-term spatial memory on the water maze, we elicited robust learning in all groups. For this, we used a more intensive training protocol (3 trials per day over 9 days). With such protocol, both control and mutant groups learned similarly well and reached minimal escape latencies after 8–9 days ($P < 0.05$ overall, Fig. 4c). Likewise, both groups rapidly learned a second platform position (not shown). Similar learning in control and mutant mice suggested here that the transgene effect was occluded by sustained training. We then tested spatial memory on several successive probe trials. One day after training, memory for the platform position was high in both groups. After 2 weeks, however, it gradually decayed in control mice ($F(1,54) = 10.861$, $P < 0.01$) and was fully extinguished by 6 weeks (Fig. 4d). But in *I-1** mutant mice, performance declined only slightly and remained significantly higher than in control mice and than baseline up to 4 weeks after training ($P < 0.05$ versus control at 2, 4 and 8 weeks), indicating that PP1 inhibition prolonged memory (Fig. 4d). We determined more precisely the temporal window of this effect by taking advantage of the inducibility of the transgene expression and inducing it either during or after training. Although *I-1** expression only during training did not prevent the weakening of memory (Fig. 4e), its induction only after learning was sufficient to maintain memory for up to 6 weeks after training ($P < 0.05$ versus control at 2 and 6 weeks, $P < 0.01$ at 4 weeks, Fig. 4f). This indicated that independent of its function during learning, PP1 specifically acts after learning to promote forgetting. It is not known whether PP1 is also involved in extinction, a form of memory erasure that, unlike forgetting, is associated with re-learning of new information.

To further investigate the physiological relevance of the PP1-dependent mechanism to forgetting, in particular in ageing-related memory decline²⁵, we tested old animals (15–18 months) in the water maze. Although both aged control and mutant mice were able to learn the platform position when intensively trained ($P < 0.05$ overall), performance was significantly lower in the control mice ($P < 0.05$ versus aged mutant, overall), suggesting a slight deficit in spatial learning (Fig. 4g). Spatial memory was also impaired in aged control mice one day after learning ($P > 0.1$ versus baseline, Fig. 4h) or later, whereas it was still robust in the mutant mice 24 h ($F(1,16) = 16.243$, $P = 0.001$ versus aged control) or even 4 weeks after learning ($P < 0.05$ versus aged control at 2 and 4 weeks, Fig. 4h), indicating again that PP1 inhibition prevents memory decline.

The present findings demonstrate the existence of a molecular constraint operated by PP1 that regulates both the acquisition and the retention of information. This constraint may preserve synaptic circuits from saturation in young individuals, but its dysregulation in old animals may precipitate memory decay^{26,27}. These data further strongly suggest that forgetting in ageing may not necessarily result from an irreversible rundown of molecular components but rather from the active intervention of PP1, possibly through mechanisms that depend on NMDA²⁸ (N-methyl-D-aspartate) receptor, opening new perspective for therapeutic treatment of

aged individuals. The possibility that *I-1** may partly mediate its effects through an unknown function independent of PP1 inhibition cannot, however, be excluded. Finally, our findings shed new light on the function of *I-1*, a PP1 inhibitor thought to be critical for synaptic plasticity²⁹; we suggest that the function of *I-1* should be further investigated by complementary electrophysiological analyses. □

Methods

Mice

We produced *tetO-I-1** mice by microinjection of the *tetO* promoter (pMM400 plasmid³⁰) fused to the *I-1** gene¹² into F₂ C57Bl/6J × CBA fertilized eggs, and backcrossed for 4–6 generations to C57Bl/6J background. CaMKII α promoter-rtTA and Cre-*lacZ* mice were also backcrossed to C57Bl/6J background. Double (*I-1*/rtTA*) and triple (*I-1*/rtTA/lacZ*) transgenic mice were obtained by heterozygous crossings. Control mice are littermates carrying no transgene or either one of the transgenes. Before testing, mice were fed dox (Westward Pharmaceuticals Corp.) at 6 mg per g food when indicated. For all experiments, control mice were treated or not treated with dox and results were pooled. Male mice 2.5–6 months old were used for behaviour. Aged males were 15–18 months old.

Assays

Polymerase chain reactions after reverse transcription of RNA (RT-PCRs) were performed as previously described⁸ using a Superscript kit (GibcoBRL) and oligonucleotides specific for the *I-1** transgene. For phosphatase assays, hippocampi and cortex were removed from adult brain and homogenized with a syringe in lysis buffer (50 mM Tris-HCl buffer at pH 7.5, 200 mM NaCl, 12 mM MgCl₂, 1 mM dithiothreitol (DTT), 0.05% NP-40, 1 mM CaCl₂) containing a cocktail of protease inhibitors (Sigma). Cellular debris were removed by ultracentrifugation (150,000g) for 1 hour at 4 °C then supernatants were desalted by gel filtration. Phosphatase activity was measured using a Biomol green assay kit (BioMol) in the presence or absence of either EGTA to block calcineurin activity, okadaic acid to block PP1 and PP2A activity, or inhibitor-2 (Bioconcept) to block PP1 activity.

Western blotting

For CREB, crude extracts were prepared from cortex by homogenization in lysis buffer (50 mM Tris-HCl pH 7.5, 20 mM EGTA, 200 mM NaCl, 12 mM MgCl₂, 1 mM DTT, 0.05% NP-40) containing a cocktail of protease and phosphatase inhibitors. For CaMKII and GluR1, membrane-enriched preparations were obtained from hippocampi by homogenization in ice-cold buffer (0.32 M sucrose, 50 mM Tris-HCl pH 7.5, 4 mM EGTA, 10 mM EDTA, 15 mM sodium pyrophosphate, 100 mM β -glycerophosphate, 10 mM NaF, 1 mM phenylmethylsulphonyl fluoride, PMSF) containing protease inhibitors, centrifugation then resuspension in buffer without sucrose. All samples were resolved by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) before electroblotting. Membranes were incubated either with anti-pSer 133 CREB (1:1,000), anti-CREB (1:1,000), anti-pThr 286 CaMKII (1:1,000), anti-CaMKII α (1:2,000), anti-pSer 845 GluR1 (1:1,000), anti-GluR1 (1:1,000) or anti- β -actin (1:1,000) antisera (Upstate Biotechnology), and visualized by chemiluminescence before densitometric quantification (NIH Image).

β -galactosidase staining

Brains were extracted immediately after testing, frozen on dry ice then kept at -80°C until used. 40- μm cryostat sagittal sections were fixed for 10 min in 2% formaldehyde and 0.2% glutaraldehyde in PBS, washed, then incubated in X-gal solution (5 mM 5-bromo-4-chloro-3-indolyl- β -D-galactoside, 5 mM K₃Fe(CN)₆, 5 mM K₄Fe(CN)₆·3H₂O, 3 mM MgCl₂) at 37 °C overnight. The number and distribution of X-gal-positive cells were determined on nine sections cut across one hemisphere (with six intervening sections) by stereology using the optical fractionator method (Stereoinvestigator software, MicroBrightfield Corp.). Sampling frames (300 × 300 × 20 μm) were used to count cells in 30–50 randomly selected areas within each tissue section.

Behaviour

The object recognition task was performed as previously described⁸ using an automated tracking system (Viewpoint). Briefly, all mice were habituated to an empty arena (63 × 51 × 25 cm) three times for 5–10 min then trained in the presence of three objects placed in fixed locations according to the protocol described in Fig. 1a. At various intervals following training, one of the familiar objects was replaced with a novel object and memory was tested in a 5-min session. Independent groups of animals were used for each time point. Discrimination ratio corresponds to the time spent exploring the novel object divided by the total time of exploration of all objects. The water maze task was performed as described previously⁸ with 60-s or 90-s training trials. Learning was evaluated by monitoring escape latencies during training with an automated tracking system (Viewpoint). For probe trials, the platform was removed from the maze and the animals were allowed a 60-s search. The same groups of mice were successively tested on each probe trial. The time (%) spent in each quadrant of the maze was recorded. Animals were left on the platform for 30 seconds after each training trial and each probe trial. Swimming speed was similar in control and mutant animals (not shown). The experimenter was blind to genotype for all behavioural tests.

Data analysis

One- or two-way analyses of variance (ANOVAs) with genotype as between-subject factor and session, block, day or quadrant as within-subject factor, followed by Fisher post-hoc tests (when necessary) were performed. Data are mean $\pm$ s.e.m.

Received 8 May; accepted 10 June 2002; doi:10.1038/nature00928.

1. Spear, N. E. *The Processing of Memories: Forgetting and Retention* (Erlbaum, Hillsdale, New Jersey, 1978).
2. Spear, N. E. & Riccio, D. C. *Memory: Phenomena and Principles* (Allyn and Bacon, Needham Heights, Massachusetts, 1994).
3. Sara, S. J. Retrieval and reconsolidation: toward a neurobiology of remembering. *Learn. Mem.* **7**, 73–84 (2000).
4. Abel, T., Martin, K. C., Bartsch, D. & Kandel, E. R. Memory suppressor genes: inhibitory constraints on the storage of long-term memory. *Science* **279**, 338–341 (1998).
5. Sweatt, J. D. Memory mechanisms: the yin and yang of protein phosphorylation. *Curr. Biol.* **11**, 391–394 (2001).
6. Mansuy, I. M., Mayford, M., Jacob, B., Kandel, E. R. & Bach, M. E. Restricted and regulated overexpression reveals calcineurin as a key component in the transition from short-term to long-term memory. *Cell* **92**, 39–49 (1998).
7. Mansuy, I. M. *et al.* Inducible and reversible gene expression with the rtTA system for the study of memory. *Neuron* **21**, 257–265 (1998).
8. Malleret, G. *et al.* Inducible and reversible enhancement of learning, memory, and long-term potentiation by genetic inhibition of calcineurin. *Cell* **104**, 675–686 (2001).
9. Lisman, J. E. A mechanism for the Hebb and anti-Hebb processes underlying learning and memory. *Proc. Natl Acad. Sci. USA* **86**, 9574–9578 (1989).
10. Morishita, W. *et al.* Regulation of synaptic strength by protein phosphatase 1. *Neuron* **32**, 1133–1148 (2001).
11. Save, E., Poucet, B., Foreman, N. & Buhot, M. C. Object exploration and reactions to spatial and nonspatial changes in hooded rats following damage to parietal cortex or hippocampal formation. *Behav. Neurosci.* **106**, 447–456 (1992).
12. Alberts, A. S., Montminy, M., Shenolikar, S. & Feramisco, J. R. Expression of a peptide inhibitor of protein phosphatase 1 increases phosphorylation and activity of CREB in NIH 3T3 fibroblasts. *Mol. Cell Biol.* **14**, 4398–4407 (1994).
13. Huang, F. L. & Glimsman, W. H. Separation and characterization of two phosphorylase phosphatase inhibitors from rabbit skeletal muscle. *Eur. J. Biochem.* **70**, 419–426 (1976).
14. Gossen, M. *et al.* Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**, 1766–1769 (1995).
15. Davis, H. P. & Squire, L. R. Protein synthesis and memory: a review. *Psychol. Bull.* **96**, 518–559 (1984).
16. Silva, A. J., Kogan, J. H., Frankland, P. W. & Kida, S. CREB and memory. *Annu. Rev. Neurosci.* **21**, 127–148 (1998).
17. Hagiwara, M. *et al.* Transcriptional attenuation following cAMP induction requires PP1-mediated dephosphorylation of CREB. *Cell* **70**, 105–113 (1992).
18. Impey, S. *et al.* Induction of CRE-mediated gene expression by stimuli that generate long-lasting LTP in area CA1 of the hippocampus. *Neuron* **16**, 973–982 (1996).
19. Gonzalez, G. A. & Montminy, M. R. Cyclic AMP stimulates somatostatin gene transcription by phosphorylation of CREB at serine 133. *Cell* **59**, 675–680 (1989).
20. Morris, R. G., Garrud, P., Rawlins, J. N. & O'Keefe, J. Place navigation impaired in rats with hippocampal lesions. *Nature* **297**, 681–683 (1982).
21. Strack, S., Barban, M. A., Wadzinski, B. E. & Colbran, R. J. Differential inactivation of postsynaptic density-associated and soluble Ca²⁺/calmodulin-dependent protein kinase II by protein phosphatases 1 and 2A. *J. Neurochem.* **68**, 2119–2128 (1997).
22. Lisman, J., Malenka, R. C., Nicoll, R. A. & Malinow, R. Learning mechanisms: the case for CaMKII. *Science* **276**, 2001–2002 (1997).
23. Lisman, J. E. & Zhabotinsky, A. M. A model of synaptic memory: a CaMKII/PP1 switch that potentiates transmission by organizing an AMPA receptor anchoring assembly. *Neuron* **31**, 191–201 (2001).
24. Blitzer, R. D. *et al.* Gating of CaMKII by cAMP-regulated protein phosphatase activity during LTP. *Science* **280**, 1940–1942 (1998).
25. Gallagher, M. & Rapp, P. R. The use of animal models to study the effects of aging on cognition. *Annu. Rev. Psychol.* **48**, 339–370 (1997).
26. Foster, T. C., Sharrow, K. M., Masse, J. R., Norris, C. M. & Kumar, A. Calcineurin links Ca²⁺ dysregulation with brain aging. *J. Neurosci.* **21**, 4066–4073 (2001).
27. Norris, C. M., Halpain, S. & Foster, T. C. Alterations in the balance of protein kinase/phosphatase activities parallel reduced synaptic strength during aging. *J. Neurophysiol.* **80**, 1567–1570 (1998).
28. Villarreal, D. M., Do, V., Haddad, E. & Derrick, B. E. NMDA receptor antagonists sustain LTP and spatial memory: active processes mediate LTP decay. *Nature Neurosci.* **5**, 48–52 (2002).
29. Connor, J. H., Quan, H., Oliver, C. & Shenolikar, S. Inhibitor-1, a regulator of protein phosphatase 1 function. *Methods Mol. Biol.* **93**, 41–58 (1998).
30. Mayford, M. *et al.* Control of memory formation through regulated expression of a CaMKII transgene. *Science* **274**, 1678–1683 (1996).

Acknowledgements

We thank S. Shenolikar for the *I-1* reagents; I. Weiss, G. Hédou, F. Dey, A. Hirschy and M. Nemir for technical help; V. Taylor for assistance with animals; A. Jongen-Relo for help with stereology; D. Benke for help with membrane-enriched preparations; and T. Bliss for reading the manuscript. This work was supported by the Swiss Federal Institute of Technology, the Swiss National Science Foundation and the National Center of Competence in Research.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.M.M. (e-mail: isabelle.mansuy@cell.biol.ethz.ch).

Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*

Peter J. Roy*, Joshua M. Stuart†, Jim Lund* & Stuart K. Kim*

* Departments of Developmental Biology and Genetics, Stanford University Medical Center, Stanford, California 94305, USA

† Stanford Medical Informatics, 251 Campus Drive, MSOB X-215, Stanford, California 94305, USA

Chromosomes are divided into domains of open chromatin, where genes have the potential to be expressed, and domains of closed chromatin, where genes are not expressed¹. Classic examples of open chromatin domains include 'puffs' on polytene chromosomes in *Drosophila* and extended loops from lampbrush chromosomes^{2,3}. If multiple genes were typically expressed together from a single open chromatin domain, the position of co-expressed genes along the chromosomes would appear clustered. To investigate whether co-expressed genes are clustered, we examined the chromosomal positions of the genes expressed in muscle of *Caenorhabditis elegans* at the first larval stage. Here we show that co-expressed genes in *C. elegans* are clustered in groups of 2–5 along the chromosomes, suggesting that expression from a chromatin domain can extend over several genes. These observations reveal a higher-order organization of the structure of the genome, in which the order of genes along the chromosome is correlated with their expression in specific tissues.

We developed a method called messenger RNA tagging to isolate muscle mRNA, because this tissue is difficult to isolate in *C. elegans*. The basis of the technique is to use a characterized promoter to express an epitope-tagged mRNA-binding protein, such as poly(A)-binding protein (PAB-1), in cells or tissues of interest (Fig. 1). Because poly(A)-binding proteins bind tightly to the poly(A) tail of mRNAs⁴, mRNAs from specific tissues can be enriched by cross-linking them to the tagged PAB-1, and co-immunoprecipitating the complex of mRNA and tagged PAB-1 using an anti-epitope monoclonal antibody²⁰. DNA microarrays can then be used to identify which mRNAs have been enriched by co-immunoprecipitation, indicating that the corresponding gene is expressed in the same cells as the tagged PAB-1.

To isolate the mRNA expressed in muscle, we first generated animals that express Flag::PAB-1 in non-pharyngeal muscles from an integrated transgene using the *myo-3* promoter⁵ (*myo-3p*; see Methods) (Fig. 2a). The mRNA–Flag::PAB-1 complex was co-immunoprecipitated from cell lysate using anti-Flag monoclonal antibodies. About 55% of the Flag::PAB-1 was immunoprecipitated from the lysate (Fig. 2b). Experiments using dot blot techniques and polymerase chain reaction with reverse transcription (RT–PCR) show that *unc-54*, which is specifically expressed in muscle⁵, was co-immunoprecipitated with the muscle-expressed Flag::PAB-1 but that *gld-1*, which is specifically expressed in the germ line⁶, was not (Fig. 2c and data not shown).

Next, we used DNA microarrays to analyse the ratio of the mRNA enriched by co-immunoprecipitation with Flag::PAB-1 relative to the mRNA present in the starting cell-free extract. Fluorescently labelled probes (see Methods) were then hybridized to DNA microarrays⁷ containing 90% of the 19,733 genes currently estimated in the *C. elegans* genome⁸. We repeated the mRNA-tagging experiment six times to assess statistically which genes are enriched.

We found that the rank order of genes that are enriched in each immunoprecipitation experiment is more consistent than their absolute level of enrichment. This indicates that the immunoprecipitation procedure enriches genes consistently relative to each

other, but that the efficiency of immunoprecipitation is variable. Hence, the percentile rank of enrichment for every gene from the six repeats was averaged together. Genes that are not enriched by mRNA tagging should have an average rank of about 50%, whereas genes expressed in muscle should have a rank significantly higher. A Student's *t*-test identified 1,364 genes that are significantly enriched in the muscle mRNA-tagging experiments ($P < 0.001$) (Fig. 3; see Supplementary Information Table 1). In control experiments using worms that do not express Flag::PAB-1 (see Methods), only 85 genes are enriched ($P < 0.001$; Supplementary Information Table 1).

We verified the mRNA-tagging approach by first showing that the list of 1,364 genes contains 16 of 18 muscle positive controls (89%) (Supplementary Information Table 2). The two positive controls that were not identified by mRNA tagging (*hlh-1* and *etr-1*) also do not show muscle expression by RNA *in situ* analysis (<http://nematode.lab.nig.ac.jp>; Supplementary Information Table 2). Second, the expression pattern for 47 of the 1,364 genes was known from prior work, and 44 are expressed in body wall muscle (94%) (Supplementary Information Table 3). In contrast, only 13 of 51 genes selected at random are expressed in muscle (25.5%; <http://nematode.lab.nig.ac.jp>; Supplementary Information Table 4). Third, we compared the list of 1,364 genes with two groups of negative controls. The first group is 559 genes that are known not to be expressed in larval stage 1 (L1) muscle. Of these, 13 are in the list of 1,364 genes (2.3%) (Supplementary Information Table 5). The second group is 7,681 genes that show undetectable hybridization signals in DNA microarray experiments using RNA from L1 hermaphrodites. Of these, only 132 are in the list of 1,364 muscle-expressed genes (1.7%) (Supplementary Information Table 6). Finally, the mRNA-tagging approach is not strongly biased against rare mRNAs. The distribution of the signal intensities of the 1,364 L1 muscle genes was found to be nearly identical to that of all genes present on the microarray (Supplementary Information Fig. 1). Taken together, these results demonstrate that the mRNA-tagging technique specifically identified a large fraction of genes expressed in *C. elegans* muscle.

We have shown that mRNA tagging is a powerful tool for profiling gene expression in specific tissues in whole-genome expression studies. There are many previously characterized promoters that can direct the expression of epitope-tagged PAB-1 in many cell types in *C. elegans*, so it is now feasible to identify mRNAs expressed in any tissue at different developmental times, under different growth conditions or in different genetic backgrounds. Because poly(A)-binding protein is highly conserved from yeast to

humans⁹, mRNA tagging is applicable to other model organisms to isolate mRNAs expressed in cells or tissues that were previously inaccessible.

The list of 1,364 muscle-expressed genes is a global overview of gene expression in *C. elegans* muscle, and provides a foundation for understanding how this tissue functions at the molecular level. Five genes encode transcription factors that may specify muscle cell function, and fourteen genes encode receptors or other synapse-associated proteins that may function at the neuromuscular junction. As expected, the list also contains housekeeping genes, such as genes involved in energy production, chromatin structure, cytoskeletal function, RNA processing and protein expression (Fig. 3c). This list also contains more than 500 genes with no known function. A more complete discussion of the list of muscle-expressed genes will be presented elsewhere.

We next looked for evidence for chromatin domains by investigating the positions of the muscle-expressed genes along the chromosomes. Specifically, we calculated how many muscle-expressed genes had start positions⁸ within 10 kilobases (kb) of the start position of another muscle gene (see Methods). This number of clustered muscle genes was then compared with the average number of clustered genes from 10,000 randomly sampled gene lists of the same size as the muscle gene list. Before the number

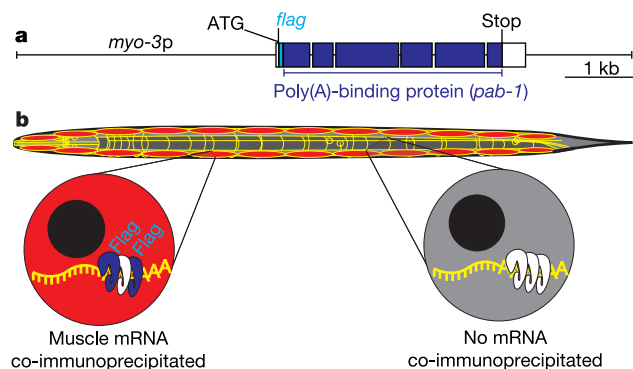


Figure 1 Outline of the mRNA-tagging technique. **a**, Schematic of *myo-3p::flag::pab-1*. The initiator methionine and stop codon are shown. Dark blue boxes show the predicted *pab-1* exons, white boxes show untranslated regions, and the light blue box represents the Flag tag. **b**, Schematic of the muscle mRNA-tagging strain. *myo-3p* drives expression of Flag::PAB-1 in non-pharyngeal muscle cells (red), but not in other tissues, such as the intestine (grey). Thus, Flag::PAB-1 (blue) binds only to muscle-expressed mRNAs, and only muscle-expressed mRNAs co-immunoprecipitate with Flag::PAB-1.

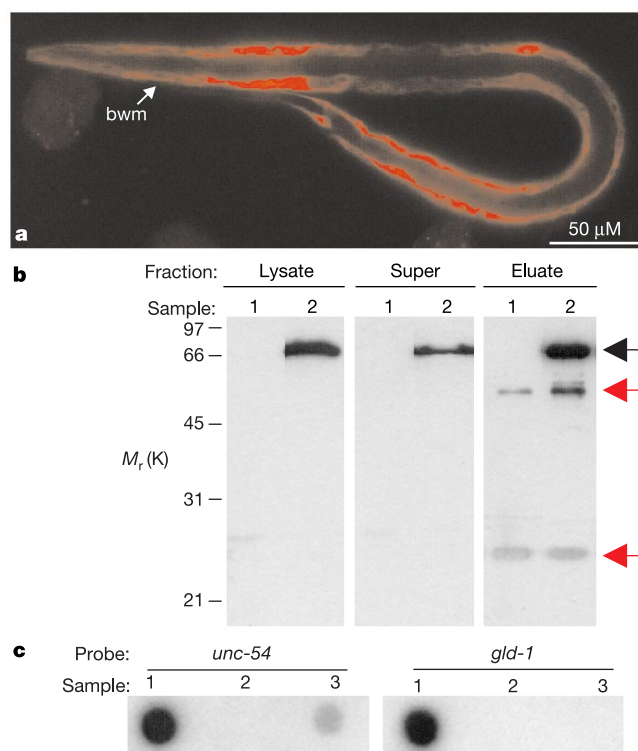


Figure 2 Muscle mRNA tagging. **a**, Anti-Flag antibody staining of a worm expressing *myo-3p::Flag::PAB-1* showing specific expression in striated muscle. bwm, body wall muscle. **b**, Western blot analysis of immunoprecipitation of Flag::PAB-1 (black arrow) from cell-free extracts. Western blots were stained with anti-Flag antibodies. Samples 1 and 2 are extracts prepared from wild-type and Flag::PAB-1-expressing animals, respectively. Lysate refers to the cell-free lysate, super refers to the supernatant after immunoprecipitation of Flag::PAB-1, and eluate refers to immunoprecipitated material. Red arrows denote the light and heavy chains of the anti-Flag antibodies. The eluate lanes contain 1.6 $\times$ more sample than the lysate and supernatant lanes. M_r , relative molecular mass. **c**, Dot blot analysis of the tissue specificity of mRNA tagging. Lane 1 is a 250-ng poly(A) mRNA standard. Lanes 2 and 3 contain RNA that was co-immunoprecipitated from wild-type and *myo-3p::Flag::PAB-1*-expressing animals, respectively. The left and right blots were hybridized with an *unc-54* probe, which is specifically expressed in body wall muscles, and a *gld-1* probe, which is expressed specifically in the germ line, respectively.

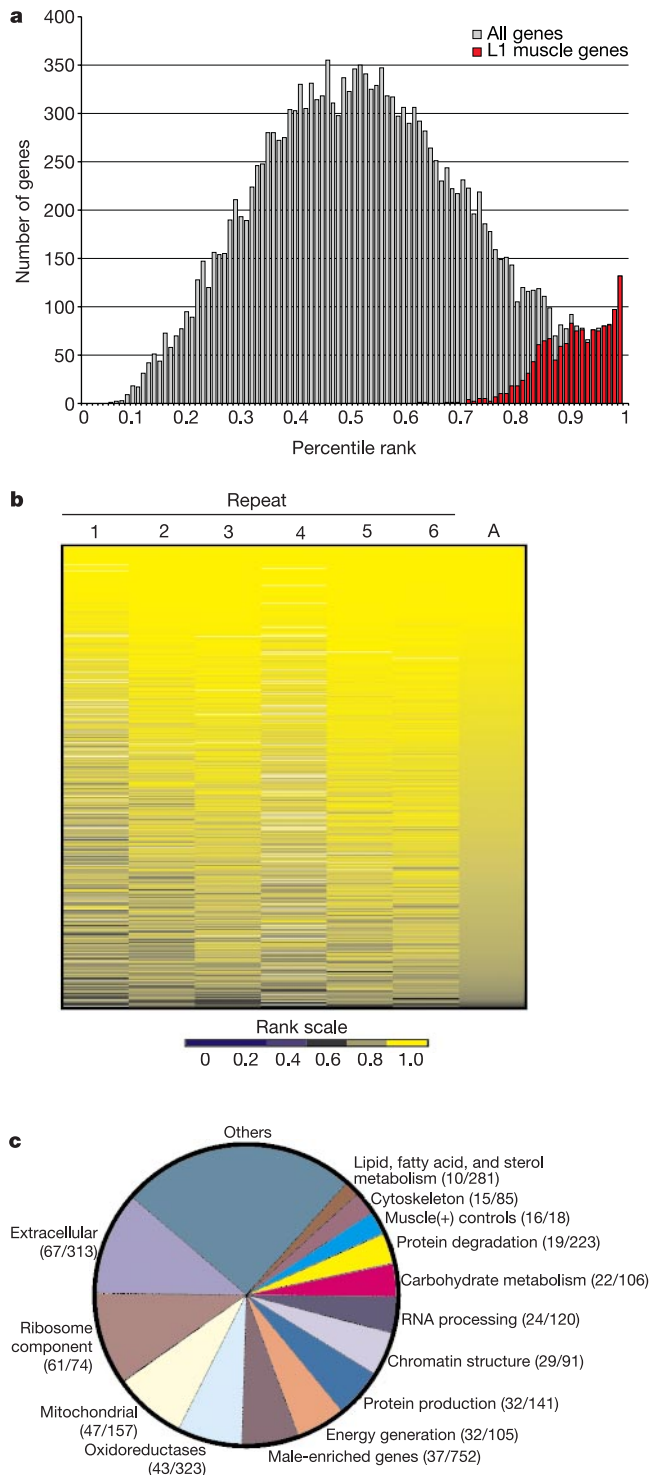


Figure 3 Genes enriched from muscle by mRNA tagging. **a**, Histogram of the average percentile rank of enrichment after mRNA tagging. The x axis shows the average percentile rank of enrichment, and the y axis shows the number of genes. Open bars represent the distribution for all genes; red bars represent the distribution of the 1,364 genes that are significantly enriched (Student's *t*-test, $P < 0.001$). **b**, Graphical representation of the percentile ranks of the 1,364 muscle-enriched genes. Each row represents a single gene. Colour represents the percentile rank of enrichment. Columns 1–6 are the six repeats and column A is the average percentile rank. Genes are listed in descending order of their average percentile rank. See Supplementary Information for how the graphical representation was generated. **c**, Pie chart showing the types of genes enriched in L1 muscle. The numbers in parentheses show the number of genes in the list that are expressed in muscle, out of the total number of genes in that biological group.

of clustered genes was calculated, however, special consideration was given to neighbouring genes that reside in the same operon, and those that may have resulted from a tandem duplication event. Operons are composed of multiple genes that are expressed from the same promoter and processed by trans-splicing to form separate mRNAs^{10,11}. Thus, clustering of different genes from the same operon on the chromosome would not be evidence for chromatin domains, and so we used only the start site of the first gene in an operon. Similarly, tandemly duplicated genes may be expressed from duplicated regulatory elements, and clustering between these genes was therefore excluded from our calculations. After removing operons and tandemly repeated genes (see Methods and Supplementary Information Figs 3 and 4), 1,304 genes remain on the list of muscle-expressed genes, and 386 of these (29.6%) are within 10 kb of another muscle-expressed gene. The number of clustered muscle-expressed genes is significantly greater than would be expected for a random distribution (310; $P < 10^{-4}$) (Table 1, Fig. 4a). We also showed that the clustering of co-expressed genes is not due to unrecognized operons or read-through transcription; we repeated this analysis considering only those neighbours that are convergently or divergently transcribed and found that they still showed significant levels of clustering ($P = 2.0 \times 10^{-6}$).

The L1 muscle genes are positioned along the chromosomes in small clusters of 2–5 genes each (Fig. 4b, c). Eighty-five of the 174 clusters that contain muscle-expressed genes (48.9%) are interrupted by a gene that is not detectably expressed. The genes in clusters contain approximately the same set of biological functions (Supplementary Information Table 7), are enriched to about the same extent by mRNA tagging, and are expressed at about the same level as the rest of the muscle-expressed genes (Supplementary Information Note 1). Hence, clustered genes have a wide variety of functions and we found no evidence that they are a specific subset of muscle-expressed genes.

We extended our analysis to determine whether other groups of genes that are expressed together exhibited clustering along the chromosomes. Previous microarray studies have identified 650 genes enriched in sperm, 258 genes enriched in oocytes, and 508 germline-intrinsic genes¹². All three germline groups showed sig-

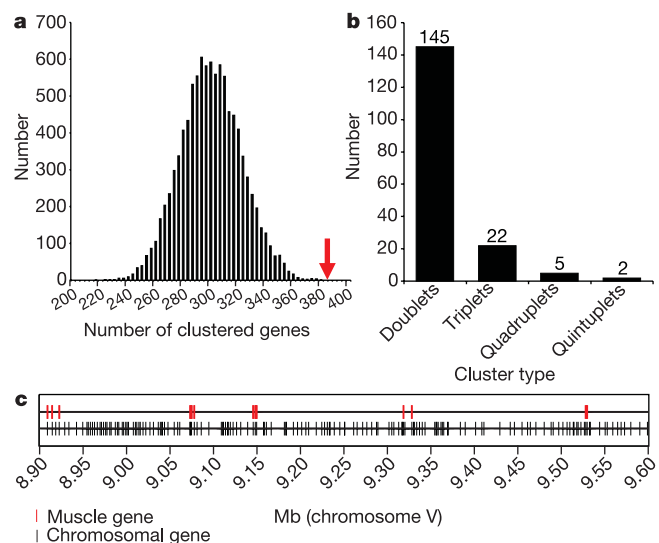


Figure 4 Muscle genes are clustered in small groups of 2–5 genes along the chromosomes. **a**, Histogram comparing 10,000 random samples with the observed muscle gene clusters (red arrow). **b**, Distribution of the number of muscle-expressed genes in each cluster. **c**, Schematic of 0.7 Mb of chromosome V, showing the positions of all genes (black), and genes expressed in muscle (red).

Table 1 Genes expressed in the same tissue tend to cluster along the chromosomes

Gene group	Total genes	Genes considered*	Observed clustered†	Expected clustered‡	$P^{\S}$
L1 muscle-enriched genes	1,364	1,304	386	310.1	7.2×10^{-5}
Sperm-enriched genes	650	616	198	112	$<10^{-15}$
Oocyte-enriched genes	258	242	26	14.4	0.017
Germline-intrinsic genes	508	475	136	67.9	$<10^{-15}$
Mountain 00	2,703	2,377	962	868.9	5.8×10^{-4}
Mountain 01	1,818	1,662	435	465.9	0.89
Mountain 02	1,465	1,100	360	270.6	$<10^{-15}$
Mountain 03	1,363	1,166	316	271.8	0.0017
Mountain 04	1,195	1,113	466	268.2	$<10^{-15}$
Mountain 05	1,012	944	248	160.7	$<10^{-15}$
Mountain 06	978	915	214	159.8	1.6×10^{-4}
Mountain 07	909	840	133	130.7	0.32
Mountain 08	810	761	225	119.8	$<10^{-15}$
Mountain 09	786	744	172	102.8	$<10^{-15}$
Mountain 10	635	555	97	64.8	0.0013
Mountain 11	587	546	133	71.5	$<10^{-15}$
Mountain 12	462	435	86	35.9	$<10^{-15}$
Mountain 13	396	368	49	27.3	1.3×10^{-5}
Mountain 14	353	331	46	22.1	4.4×10^{-5}
Oxidoreductases	323	317	21	16.7	0.23
Transcription factors	255	214	8	8.6	0.56
Lipid, fatty acid and sterol metabolism genes	281	280	12	12.4	0.53

The genes in each group, except for the mountains, are given in Supplementary Information Table 1. The genes in each mountain can be accessed elsewhere (http://cmgm.stanford.edu/~kimlab/topomap/c_elegans_topomap.htm).

*The number of genes considered, after removing operons, putative tandem duplicates, and genes with uncertain genomic positions.

†The number of genes in the gene group that are within 10 kb of another gene from the same group.

‡The number of clusters expected for a random distribution. Shown are the average number of clustered genes from 10,000 repetitions of randomly selecting the same number of genes from the genome as the experimental gene group and counting how many genes are within 10 kb of another gene from the same randomly selected list.

§The probability that the observed number of genes that cluster could be matched or exceeded by chance.

nificant levels of gene clustering along the chromosomes ($P < 0.05$) (Table 1). A different way to find genes that are expressed together is to use expression profiles from a large number of microarray experiments. Specifically, 553 diverse microarray experiments were used to define 44 sets of co-regulated genes reflecting similar expression during development, under different growth conditions and in many different types of mutants¹³. In all, 13 of the 15 largest gene clusters showed significant levels of gene clustering ($P < 0.05$) (Table 1).

As a negative control, we analysed gene clustering using three groups of genes that are predicted to be expressed in a variety of different cell types at different times in development: genes encoding oxidoreductases, transcription factors and proteins involved in lipid, fatty acid and sterol metabolism (Table 1). None of these groups of genes is clustered along the chromosomes ($P > 0.2$).

Overall, we have found compelling evidence that genes expressed together in the same tissues or co-regulated in diverse microarray experiments are clustered in small groups along the chromosomes. Recent work has shown that genes that are highly expressed in a variety of human tissues or that are co-expressed in yeast are clustered along the chromosomes^{14,15}. These results suggest that gene clustering of functionally related genes may occur in all metazoans.

One interpretation of the results presented here is that gene clusters may correspond to regions of active chromatin. Because opening chromatin for one gene can result in the opening of neighbouring chromatin^{16,17}, transcriptional machinery may access two co-expressed genes more efficiently if they are neighbours than if they are far apart. In this model, the fortuitous juxtaposition of co-expressed genes would be selected over an evolutionary time-scale. An alternative interpretation of clustering is that genes expressed in the same tissue appear in clusters because neighbouring genes share a single DNA enhancer element. In this model, neighbouring genes appear clustered depending on the strength of nearby enhancers and would not require genomic rearrangements. In conclusion, our observations that co-expressed genes are clustered along the chromosomes of *C. elegans* provide direct evidence that there is a higher-order organization of genes within the genome. □

Methods

Molecular biology

Standard techniques of molecular biology and worm culture were used to respectively generate the *myo-3p::flag::pab-1* construct (pPRSK9) and the resulting SD1075 strain (see Supplementary Information for details).

Messenger RNA tagging

The mRNA-tagging protocol is a modification of a formerly published protocol¹⁸. For each sample, 1 ml of packed worms was resuspended in M9 buffer, and then fixed in 0.5% formaldehyde in M9 for 1 h at 4 °C. Animals were rinsed in homogenization buffer (HB), and then resuspended in 3 ml of HB. HB is prepared by diethyl pyrocarbonate (DEPC) treatment of 300 mM NaCl, 50 mM Hepes buffer at pH 7.6, 10 mM MgCl₂, 1 mM EGTA, 30 mM EDTA, 0.2 mg ml⁻¹ heparin (sodium salt) and 10% glycerol, and then adding dithiothreitol (DTT) to 1 mM, vanadyl ribonucleoside complex to 8 mM (Sigma), 50 U ml⁻¹ RNasin, and half a protease inhibitor cocktail tablet (Roche). Animals were lysed with two or three passes on a French press at 8,000 p.s.i. (~55 MPa), and then 25 passes on a Wheaton homogenizer. Large debris was sedimented by centrifugation at 4,300g for 6 min, and then smaller debris was separated from the supernatant by centrifugation at 48,000g for 20 min. At this point, the lysate could be flash frozen and stored for future use.

RNA bound to Flag::PAB-1 was enriched with anti-Flag-M2 affinity gel beads (Sigma). The affinity beads were prepared by rinsing twice in RNase-free glycine HCl (pH 3.5) and then four times in HB at 4 °C. The beads were collected by centrifugation, and stored at 4 °C. We mixed together 1 ml of lysate, 50 U of RNasin, 8 µl of 200 mM vanadyl ribonucleoside, and 75 µl of pelleted anti-Flag-M2 affinity gel beads for 2–3 h at 4 °C. The affinity beads were washed four times with cold HB, and then the RNA–protein crosslinks were reversed by incubating the beads in 125 µl of elution buffer (50 mM Tris/HCl at pH 8.0, 10 mM EDTA, 1.3% SDS; 20 U RNasin) at 65 °C for 30 min. We collected the supernatant containing the RNA, repeated the elution once more, and then combined the two supernatants.

RNA was isolated by mixing 250 µl of the immunoprecipitation supernatant (or cell extract) with 1 ml of trizol, and then mixing 250 µl of chloroform. After letting the mixture stand for 10 min, it was spun in a microfuge at 12,000 g for 15 min at 4 °C, and then the supernatant was extracted with chloroform. RNA was precipitated by adding 600 µl of isopropanol, incubating at room temperature for 10 min and then centrifuging in a microfuge for 20 min at 12,000 g at 4 °C. The RNA pellet was rinsed twice with 1 ml of 70% ethanol.

RNA was linearly amplified as previously described¹⁹. The DNA microarrays, probe preparation, and microarray hybridization are described in Supplementary Information.

Data analysis (mRNA tagging)

To identify genes that are significantly enriched by mRNA tagging, we first normalized the total amount of cy3 and cy5 signal to each other in each hybridization. We measured the ratio of the signals from the co-immunoprecipitated RNA (Cy-5) to total RNA in the cell extract (Cy-3), and then calculated the percentile rank for each gene relative to all genes in each hybridization. The mRNA-tagging experiment was repeated six times, and the mean percentile rank from all repeats was determined. A Student's *t*-test was used to determine which genes showed a mean enrichment significantly greater than the mean enrichment for all genes. Mock mRNA tagging was done using four repeats with wild-type (N2) worms.

Data analysis (chromosomal cluster analysis)

A gene from the muscle gene list was counted as clustered if its start position was within 10 kb of the start position of another muscle gene. We also varied the distance criteria between 1 kb and 1 Mb and observed significant clustering ($P < 0.001$) from 1 to 25 kb.

A detailed explanation of the calculations used to measure gene clustering is given in Supplementary Information. Briefly, the calculations included only those genes that were present on the microarray and for which we could determine a chromosomal position⁸. The calculations used only the first gene in an operon, and only one gene of tandem repeats. The number of clusters expected for a random distribution was calculated separately for each chromosome, and then summed to give the total number. This was done so that the number expected for a random distribution reflected any bias in the experimental list. Finally, because the germline data were obtained from experiments using microarrays containing 11,917 genes¹², lists of genes were randomly selected from only these genes to avoid bias.

Received 23 January; accepted 16 July 2002; doi:10.1038/nature01012.

1. Weintraub, H. Tissue-specific gene expression and chromatin structure. *Harvey Lect.* **79**, 217–244 (1984).
2. Beermann, W. Control of differentiation at the chromosomal level. *J. Exp. Zool.* **157**, 49–62 (1964).
3. Gall, J. G. & Callan, H. G. H3 Uridine incorporation in lampbrush chromosomes. *Proc. Natl Acad. Sci. USA* **48**, 562–570 (1962).
4. Grolach, M., Burd, C. G. & Dreyfuss, G. The mRNA poly(A)-binding protein: localization, abundance, and RNA-binding specificity. *Exp. Cell Res.* **211**, 400–407 (1994).
5. Okkema, P. G., Harrison, S. W., Plunger, V., Aryana, A. & Fire, A. Sequence requirements for myosin gene expression and regulation in *Caenorhabditis elegans*. *Genetics* **135**, 385–404 (1993).
6. Jones, A. R., Francis, R. & Schedl, T. GLD-1, a cytoplasmic protein essential for oocyte differentiation, shows stage- and sex-specific expression during *Caenorhabditis elegans* germline development. *Dev. Biol.* **180**, 165–183 (1996).
7. Jiang, M. *et al.* Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **98**, 218–223 (2001).
8. Stein, L., Sternberg, P., Durbin, R., Thierry-Mieg, J. & Spieth, J. WormBase: network access to the genome and biology of *Caenorhabditis elegans*. *Nucleic Acids Res.* **29**, 82–86 (2001).
9. Fukami-Kobayashi, K., Tomoda, S. & Go, M. Evolutionary clustering and functional similarity of RNA-binding proteins. *FEBS Lett.* **335**, 289–293 (1993).
10. Blumenthal, T. Trans-splicing and polycistronic transcription in *Caenorhabditis elegans*. *Trends Genet.* **11**, 132–136 (1995).
11. Blumenthal, T. Gene clusters and polycistronic transcription in eukaryotes. *BioEssays* **20**, 480–487 (1998).
12. Reinke, V. *et al.* A global profile of germline gene expression in *C. elegans*. *Mol. Cell* **6**, 605–616 (2000).
13. Kim, S. K. *et al.* A gene expression map for *Caenorhabditis elegans*. *Science* **293**, 2087–2092 (2001).
14. Cohen, B. A., Mitra, R. D., Hughes, J. D. & Church, G. M. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nature Genet.* **26**, 183–186 (2000).
15. Caron, H. *et al.* The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **291**, 1289–1292 (2001).
16. Hebbes, T. R., Clayton, A. L., Thorne, A. W. & Crane-Robinson, C. Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken β -globin chromosomal domain. *EMBO J.* **13**, 1823–1830 (1994).
17. Stalder, J. *et al.* Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNase I. *Cell* **20**, 451–460 (1980).
18. Goodwin, E. B., Okkema, P. G., Evans, T. C. & Kimble, J. Translational regulation of tra-2 by its 3' untranslated region controls sexual identity in *C. elegans*. *Cell* **75**, 329–339 (1993).
19. Wang, E., Miller, L. D., Ohnmacht, G. A., Liu, E. T. & Marincola, F. M. High-fidelity mRNA amplification for gene profiling. *Nature Biotechnol.* **18**, 457–459 (2000).
20. Tenenbaum, S. A., Carson, C. C., Lager, P. J. & Keene, J. D. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc. Natl Acad. Sci. USA* **97**, 14085–14090 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank J. Shaw, K. Mach, M. Laub, J. Lieb, M. Hiller, E. Harmon and J. Wang for discussions; T. Blumenthal and J. Wang for sharing unpublished results; J. Wang for wild-type poly(A) mRNA; and M. Kiraly for assistance with the Supplementary Information. We also thank the programmers at the Stanford Microarray Database for microarray analysis and database management, and Proteome and Wormbase for annotation of *C. elegans* genes. P.J.R. is a Stanford University Beckman Fellow. This work was supported by a Human Frontiers Fellowship (P.J.R.) and grants from the National Institutes of Health (S.K.K.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.K.K. (e-mail: kim@cmgm.stanford.edu).

Shh and Gli3 are dispensable for limb skeleton formation but regulate digit number and identity

Ying Litington*†, Randall D. Dahn†‡, Yina Li*, John F. Fallon‡ & Chin Chiang*

* Vanderbilt University Medical Center, Department of Cell and Developmental Biology, Nashville, Tennessee 37232, USA

‡ Department of Anatomy, University of Wisconsin, 1300 University Avenue, Madison, Wisconsin 53706, USA

† These authors contributed equally to this work

Most current models propose *Sonic hedgehog* (*Shh*) as the primary determinant of anteroposterior development of amniote limbs¹. *Shh* protein is said to be required to direct the formation of skeletal elements and to specify digit identity through dose-dependent activation of target gene expression. However, the identity of genes targeted by *Shh*, and the regulatory mechanisms controlling their expression, remain poorly understood. *Gli3* (the gene implicated in human Greig cephalopolysyndactyly syndrome) is proposed to negatively regulate *Shh* by restricting its expression and influence to the posterior mesoderm^{2–4}. Here we report genetic analyses in mice showing that *Shh* and *Gli3* are dispensable for formation of limb skeletal elements: *Shh*^{−/−} *Gli3*^{−/−} limbs are distally complete and polydactylous, but completely lack wild-type digit identities. We show that the effects of *Shh* signalling on skeletal patterning and ridge maintenance are necessarily mediated through *Gli3*. We propose that the function of *Shh* and *Gli3* in limb skeletal patterning is limited to refining autopodial morphology, imposing pentadactyl constraint on the limb's polydactyl potential, and organizing digit identity specification, by regulating the relative balance of *Gli3* transcriptional activator and repressor activities.

In *Drosophila*, signalling by Hedgehog (Hh) is necessarily mediated through the regulation of opposing activator and repressor activities of the bifunctional transcription factor cubitus interruptus (Ci)^{5,6}. Ci undergoes default proteolysis to truncated forms (CiR) that repress expression of *hh* target genes⁷; stimulation of Hh prevents CiR formation and generates a full-length transcriptional activator (CiA) that upregulates expression of *hh* target genes⁵. *Shh* signalling in vertebrates is mediated by the GLI family (*Gli1–3*) of Ci orthologues, and evidence suggests that the activator and repressor functions of Ci have been uncoupled and partitioned among distinct GLI proteins¹.

Anteroposterior (A/P) patterning in the limbs of amniotes is controlled by posterior limb bud mesoderm, called the zone of polarizing activity (ZPA), through secretion of *Shh* protein; ectopic *Shh* or ZPA tissue grafted to the anterior border induces mirror-image digit duplications⁸. Gain- and loss-of-function studies suggest that *Shh* signalling progressively specifies increasing numbers of digits, with more-posterior identities, through dose-dependent activation of putative *Shh* target genes^{9–13}, but the identity and function of *Shh* effector genes remain unclear¹. *Gli1* has been proposed to positively mediate *Shh* signalling (reviewed in ref. 1), but analysis of single- and double-mutant combinations of *Gli1–3* demonstrates that *Gli1* and *Gli2* perform no significant limb A/P patterning function¹. Studies *in vivo* and *in vitro* suggest that *Gli3*, normally expressed in an anterior domain complementary to *Shh*³, negatively regulates the expression of both *Shh* and its target genes through generating a repressor form (*Gli3R*) in a manner analogous to that in *Drosophila*^{2,4,14}. *Xenopus*^{15,16} and cell culture studies provide evidence that *Gli3* may positively mediate *Shh* signalling in the limb; *Shh* stimulation prevents *Gli3R* formation^{4,14}, and *Gli3*

can form activator species (Gli3A) that upregulate Shh-responsive gene expression^{14,17,18}.

We examined the genetic interactions of *Shh* and *Gli3* to understand their function in A/P patterning of mouse limbs. *Shh*^{-/-} limbs have severe distal A/P skeletal deficiencies; all zeugopod and autopod elements are either missing, fused or lack normal identity, except for an identifiable digit 1 (d1) in the hindlimb^{10,11} (Fig. 1b). The semidominant mouse *Extra toes* (*Xt*¹) mutation generates a *Gli3* null allele resulting in preaxial d1 iterations in heterozygotes¹⁹. The limbs of *Xt*¹/*Xt*¹ homozygotes (hereafter *Gli3*^{-/-}) are fairly normal to the level of the wrist/ankle, but exhibit severe polydactyly¹⁹ (Fig. 1c), which has been causally attributed to ectopic *Shh* expression in early-stage mesoderm of the anterior limb bud^{2,3}.

Shh^{-/-} *Gli3*^{-/-} mice develop distally complete and polydactylous limbs (Fig. 1d). These limbs develop virtually normal stylopod and zeugopod elements, but have 6–11 syndactylous digits per limb. *Shh*^{-/-} *Gli3*^{-/-} digits exhibit complete loss of wild-type identities (see Fig. 1 legend; digit identity is defined in Supplementary Information). These genetic data demonstrate that *Shh* and *Gli3* are not required to elaborate formation of limb skeletal elements

along the A/P axis; on the contrary, they act to constrain the polydactylous potential of the autopod. We conclude that the requirement for *Shh* and *Gli3* in normal patterning of the limb skeleton is restricted to modifying two aspects of autopod morphology: imposing pentadactyl constraint and specifying normal digit identities.

We were interested to note that *Shh*^{-/-} *Gli3*^{-/-} and *Gli3*^{-/-} limbs are virtually indistinguishable (compare Fig. 1d and c), despite the endogenous and ectopic *Shh* expression in *Gli3*^{-/-} limb buds^{2,3,20}. *Shh* therefore has no effect on skeletal patterning in the absence of *Gli3*. *Gli3*^{-/-} polydactyly must be a direct effect of losing *Gli3* function, and is not due to ectopic expression of *Shh* or *dHand* (see below)^{3,21}. This observation explains why *Gli3*^{-/-} limbs clearly lack the mirror-image digit duplications expected from an ectopic anterior ZPA. Our analysis thus implicates *Gli3* as an obligate component of ZPA function, required in responding cells for Shh to exert polarizing activity. Despite the polydactylous potential of *Shh*^{-/-} *Gli3*^{-/-} limb bud mesoderm, grafts to chick hosts do not induce supernumerary elements (data not shown), confirming that *Shh* is required for ZPA activity.

The development of *Shh*^{-/-} *Gli3*^{+/-} mice revealed that, in the absence of *Shh*, removing *Gli3* restores limb skeletal elements in a dose-dependent manner (Fig. 1e). *Shh*^{-/-} *Gli3*^{+/-} limbs develop two identifiable zeugopod elements with clear A/P asymmetry. *Shh*^{-/-} *Gli3*^{+/-} limbs typically form three or four digits, inter-

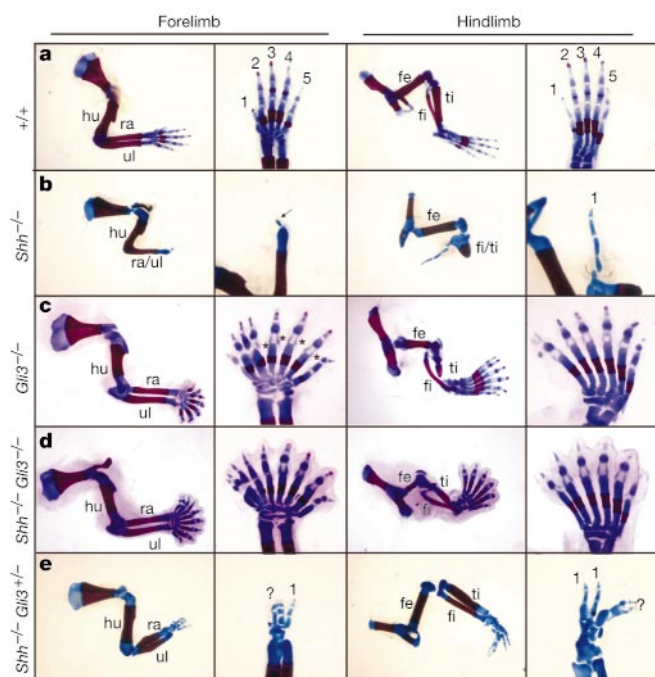


Figure 1 *Gli3* regulates digit number and identity. Skeletal analysis of E16.5 wild-type (a), *Shh*^{-/-} (b), *Gli3*^{-/-} (c), *Shh*^{-/-} *Gli3*^{-/-} (d) and *Shh*^{-/-} *Gli3*^{+/-} (e) limbs. Posterior is to the bottom (first and third columns) or right (second and fourth columns). a, Wild-type stylopod (hu, humerus; fe, femur), zeugopod (ra, radius; ul, ulna; ti, tibia; fi, fibula) and autopod (digits 1–5) elements are labelled. b, *Shh*^{-/-} stylopods are normal and zeugopods are reduced or fused. Autopods have a single unidentifiable element in the forelimb (arrow) and a digit 1 in the hindlimb. c, *Gli3*^{-/-} stylopod and zeugopod elements are normal except for shortened femurs and variable tibial aplasia, and autopods are polydactylous. d, *Shh*^{-/-} *Gli3*^{-/-} limbs are normal to the wrist or ankle, but are polydactylous, and digits lack normal identities. The syndactylous digits are short and thick, and appear more serially homologous than in the wild type; phalanges exhibit alternating gaps of unstained cartilage and bulbous stained material (see asterisks in c). *Shh*^{-/-} *Gli3*^{-/-} autopods exhibit reduced, but characteristic, A/P polarity; note the invariant bunching of anterior digits and kinking of the posteriormost forelimb digit. *Shh*^{-/-} *Gli3*^{-/-} and *Gli3*^{-/-} limbs are virtually indistinguishable (compare d with c); minor differences are within each genotype's phenotypic range. e, *Shh*^{-/-} *Gli3*^{+/-} zeugopod elements are distinguishable as tibia/fibula and radius/ulna. *Shh*^{-/-} *Gli3*^{+/-} autopods develop three or four digits; all unfused digits are identifiable as digit 1, regardless of A/P position.

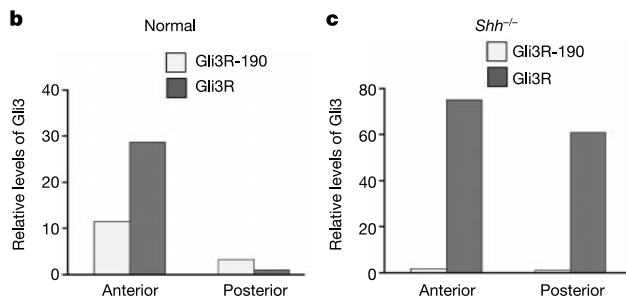
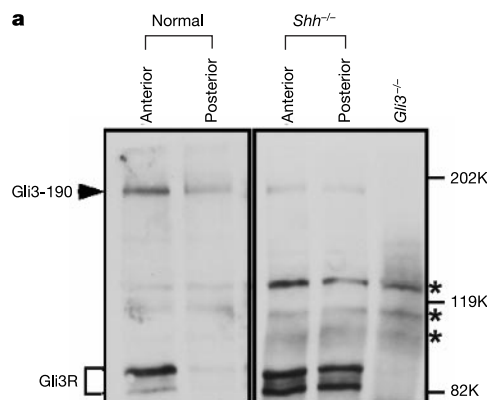


Figure 2 Shh signalling regulates Gli3 processing in the mouse limb. a, Protein extracts from E10.5 wild-type, *Shh*^{-/-} and *Gli3*^{-/-} limb buds were immunoblotted and incubated with Gli3-specific antibodies recognizing a prominent 190K band and several 83–86K bands, corresponding to full-length (Gli3-190) and processed repressor forms (Gli3R) of Gli3, respectively. b, c, Histograms showing Gli3-190 and Gli3R protein levels in anterior and posterior halves of wild-type (b) and *Shh*^{-/-} (c) limb buds. Absolute Gli3-190 and Gli3R levels both form decreasing anterior-to-posterior gradients in wild-type limbs (b), but relative levels of Gli3-190:Gli3R form an increasing anterior-to-posterior gradient (0.4 in the anterior, 3.3 in the posterior; data not shown). *Shh*^{-/-} limbs generate uniformly high levels of Gli3R and negligible Gli3-190 (c). Levels of Gli3 protein were normalized to Gli3R (b) and Gli3-190 (c) levels in posterior wild-type and *Shh*^{-/-} limb buds, respectively. Gli3 antibodies recognize several non-specific proteins (asterisks) that are present in *Gli3*^{-/-} limb buds.

mediate to the single digit in *Shh*^{-/-} and the 6–11 digits in *Shh*^{-/-} *Gli3*^{-/-} limbs. The direct correlation between *Gli3* levels and digit number supports the hypothesis that *Gli3*, rather than *Shh*, directly regulates digit number. Also, all unfused *Shh*^{-/-} *Gli3*^{+/-} digits are identifiable as d1 regardless of position along the A/P axis. Therefore, although reducing *Gli3* dosage in *Shh*^{-/-} limbs can mimic Shh signalling in rescuing digit formation, posterior digit identities are not restored; the regulation of digit number and identity are *Gli3*-dependent, but separable, processes.

Shh regulates limb skeletal patterning by modulating Gli3 activity, so we examined Gli3 processing in anterior and posterior halves of wild-type and mutant limb buds (Fig. 2). Wild-type limbs generate a decreasing anterior-to-posterior gradient of Gli3 repressor (Gli3R) (ref. 4; Fig. 2a, b). Polarized distribution of Gli3R is abolished in *Shh*^{-/-} mutants; high Gli3R levels are detectable throughout *Shh*^{-/-} limb bud mesoderm (Fig. 2a, c). Our results are consistent with Shh signalling preventing Gli3R formation and generating asymmetric levels of Gli3R across the A/P axis⁴. Our data further suggest that deficiencies of the *Shh*^{-/-} limb skeleton are caused by uniform, high-level production of Gli3R throughout the limb bud mesoderm. Studies of the neural tube²² support the possibility that severe *Shh*^{-/-} phenotypes observed in other tissues may also be due to negative effects of uniform Gli3R; analysis of *Shh*^{-/-} *Gli3*^{-/-} mutants would distinguish these effects from the loss of 'positive' Shh signalling input.

Increasing ratios of full-length Gli3-190:Gli3R form an increasing anterior-to-posterior gradient. Relative Gli3-190:Gli3R levels differ eightfold between wild-type anterior (0.4) and posterior (3.3) halves (Fig. 2a, b). Relative Gli3-190:Gli3R levels in *Shh*^{-/-} anterior and posterior limb bud halves (0.02) are roughly equivalent (Fig. 2a, c). Because *Shh*^{-/-} limbs can form normal d1s, we assumed that

Gli3-190:Gli3R levels in *Shh*^{-/-} limb buds correlate to ratios present in wild-type anterior limb regions exposed to little, if any, direct Shh signalling¹⁰. Therefore, our analysis suggests that Gli3-190:Gli3R levels vary at least 165-fold across the limb A/P axis, whereas absolute Gli3R levels vary only 65-fold (data not shown). Evidence from *in vivo*^{15,16} and *in vitro*^{14,17,18} studies shows that full-length Gli3 can function as a transcriptional activator (Gli3A) to positively mediate Shh signalling. We propose that limb bud Gli3-190 represents Gli3A, and that the balance of Gli3A:Gli3R transcriptional activities, rather than Gli3R levels alone⁴, modulate expression of Shh target genes and control autopod patterning. It is notable that the balance of Gli3A:Gli3R activities would offer a significant gain in regulatory potential (>165-fold versus 65-fold) over Gli3R levels alone, and would directly parallel the mediation of Hh signalling by Ci (orthologues of *Shh* and *Gli3*, respectively) in *Drosophila* embryogenesis^{5,6}.

Our skeletal analyses reveal that limb mesoderm possesses previously unappreciated and *Shh*-independent morphogenetic potential, so we next assessed the expression of putative *Shh* target genes. Expression of wild-type *Ptch1* and *Gli1* is posteriorly restricted (ref. 8; Fig. 3a, f) and coincident with Shh protein localization¹⁰. In contrast, the limbs of all *Shh*^{-/-} genotypes (*Shh*^{-/-}, *Shh*^{-/-} *Gli3*^{-/-} and *Shh*^{-/-} *Gli3*^{+/-}) did not detectably express *Gli1* or *Ptch1* (Fig. 3b, d, e, g, i, j). In *Gli3*^{-/-} limbs, however, spatial domains of *Gli1* and *Ptch1* expression are unaffected (Fig. 3c, h) but *Gli1* is downregulated, suggesting that normal expression levels require positive *Gli3* input. Importantly, *Gli1* and *Ptch1* expression verifies that Shh signalling is intact in *Gli3*^{-/-} limbs despite having no effect on skeletal pattern. These data confirm that *Gli1* and *Ptch1* are *Shh*-responsive genes whose expression reflects high-level Shh signalling⁸, and demonstrate that *Shh*^{-/-} *Gli3*^{-/-} polydactyly is

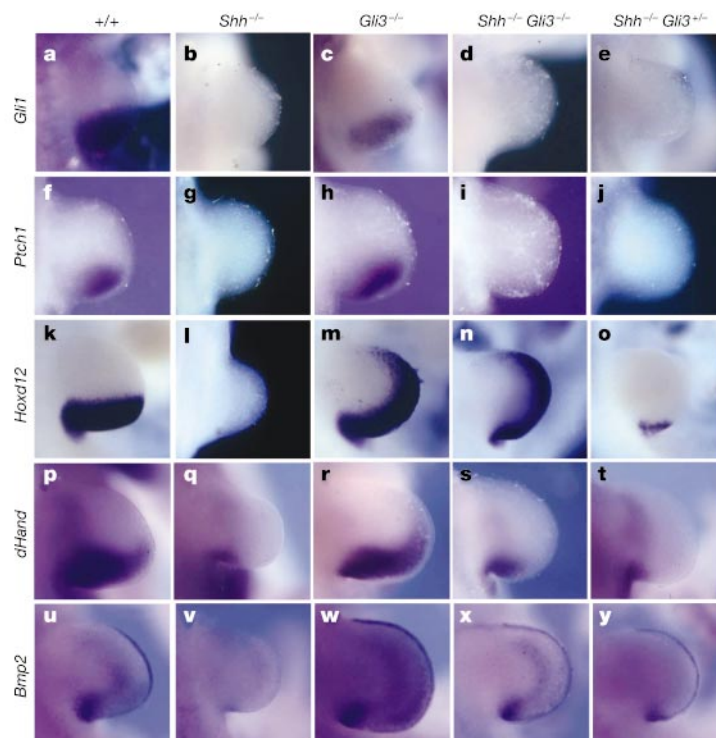
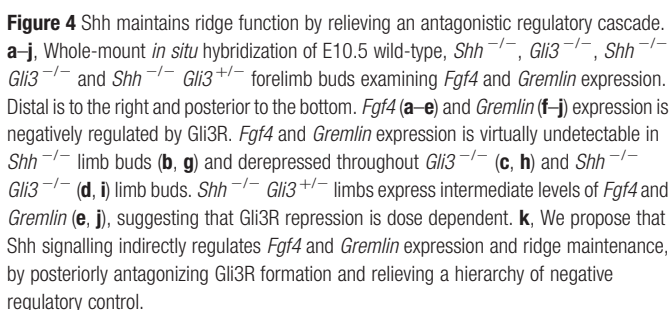


Figure 3 Regulation of putative SHH target genes in the limb. E10.5 wild-type, *Shh*^{-/-}, *Gli3*^{-/-}, *Shh*^{-/-} *Gli3*^{-/-} and *Shh*^{-/-} *Gli3*^{+/-} forelimb buds were examined for expression of *Gli1*, *Ptch1*, *Hoxd12*, *dHand* and *Bmp2*. Distal is to the right and posterior to the bottom. **a–j**, Posterior *Gli1* (**a–e**) and *Ptch1* (**f–j**) expression requires *Shh*, independently of *Gli3*. **k–o**, *Hoxd12* expression is negatively regulated by Gli3R. *Hoxd12* is undetectable in *Shh*^{-/-} limb buds (**l**), but is derepressed across the A/P axis in *Gli3*^{-/-}

(**m**) and *Shh*^{-/-} *Gli3*^{-/-} (**n**) limb buds, irrespective of continued Shh signalling in *Gli3*^{-/-} limbs. *Hoxd12* derepression is posteriorly restricted in *Shh*^{-/-} *Gli3*^{+/-} limbs (**o**). **p–t**, Gli3R is not required to posteriorly restrict *dHand* expression. Low-level *dHand* expression is detected in anterior *Gli3*^{-/-} limb mesoderm (**r**), but is posteriorly restricted in *Shh*^{-/-} *Gli3*^{-/-} limbs (**s**). **u–y**, *Bmp2* expression is spatially similar in all genotypes, but levels are reduced in *Shh*^{-/-} limbs (**v**).

Proximodistal limb outgrowth requires sustained fibroblast growth factor gene (*Fgf*) signals from the apical ridge at the distal margin of the limb bud²⁴. Current hypotheses propose that *Shh* maintains ridge function by inducing *Gremlin*, the antagonist against Bmp, neutralizing the Bmp activity that negatively regulates ridge function²⁵. *Fgf4* expression is normally restricted to the posterior two-thirds of the apical ridge (ref. 24; Fig. 4a) and is not maintained in *Shh*^{-/-} apical ridges^{11,12,25,26} (Fig. 4b), suggesting that

Our results further suggest that two distinct classes (groups 1 and 2) of molecular mechanisms underlie polydactyly, which is one of the most common human congenital malformations²⁸. Limb buds of group 1 mutants, such as *Alx4*^{-/-} mice²⁹, express ectopic anterior *Shh* without disrupting *Gli3* function, thus initiating an ectopic anterior ZPA. Group 1 limbs develop non-syndactylous, preaxial mirror-image digit duplications that possess normal identities²⁹. In contrast, group 2 mutations disrupt normal Gli3 function in a *Shh*-independent manner. One of several examples of group 2 mutants is *ta*². GLI3 hemizyosity causes preaxial digit 1 iterations in *Gli3*^{+/-} mice and human Grieg cephalopolysyndactyly (GCPs) syndrome^{19,30}, whereas *Gli3*^{-/-} mice have generalized (not pre-axial) polydactyly and lack normal digit identity. Truncated GLI3 proteins



with aberrant transcriptional properties¹⁷ cause semidominant human Pallister–Hall (PHS) and post-axial polydactyly (PAP-A) syndromes, which respectively result in medial and post-axial polydactyly^{17,30}. Semidominant group 2 mutant alleles function in the context of intact Shh signalling and a normal *Gli3* allele, and thus group 2 mutations collectively present a heterogeneous spectrum of phenotypes. However, group 2 phenotypes are characterized by a lack of mirror-image digit duplications, and affected digits are nearly always syndactylous. Therefore, the position and identity (or lack thereof) of supernumerary digits resulting from different group 2 mutations may vary, and presumably reflects how particular mutations alter the transcriptional output of Gli3-regulated target genes. □

Methods

Animals

The generation and identification of *Shh* and *Gli3* mutant mice were performed as described²².

Staining and whole-mount *in situ* hybridization

Alcian blue and alizarin red staining of cartilage and bone, and whole mount *in situ* hybridization analyses, were performed as described²².

Antibody production and western analysis

Amino-terminal-specific GLI3 antibodies were generated against amino acids 396–500 of human GLI3. Fusion proteins of GLI3 and glutathione S-transferase (GST) were expressed by *Escherichia coli*, isolated and used to immunize rabbits, according to standard protocols. Antiserum was affinity purified using GST–GLI3 fusion proteins immobilized on Affi-gel 10 beads (Bio-Rad). We collected 150 µg of protein lysate samples from anterior and posterior limb bud halves, and resolved them on 7.5% SDS–polyacrylamide gels. Using anti-N-terminal GLI3 (1:400) primary antibodies and biotinylated anti-rabbit immunoglobulin-γ secondary antibodies, we visualized protein bands by incubation with a streptavidin-peroxidase conjugate followed by an enhanced chemiluminescence detection method (Amersham).

Received 12 April; accepted 2 August 2002; doi:10.1038/nature01033.

Published online 18 August 2002.

- Ingham, P. W. & McMahon, A. P. Hedgehog signaling in animal development: paradigms and principles. *Genes Dev.* **15**, 3059–3087 (2001).
- Masuya, H., Sagai, T., Wakana, S., Moriawaki, K. & Shiroishi, T. A duplicated zone of polarizing activity in polydactylous mouse mutants. *Genes Dev.* **9**, 1645–1653 (1995).
- Büscher, D., Bosse, B., Heymer, J. & Rüther, U. Evidence for genetic control of *Sonic hedgehog* by *Gli3* in mouse limb development. *Mech. Dev.* **62**, 175–182 (1997).
- Wang, B., Fallon, J. F. & Beachy, P. A. Hedgehog-regulated processing of Gli3 produces an anterior/posterior repressor gradient in the developing vertebrate limb. *Cell* **100**, 423–434 (2000).
- Méthot, N. & Basler, K. Hedgehog controls limb development by regulating the activities of distinct transcriptional activator and repressor forms of *Cubitus interruptus*. *Cell* **96**, 819–831 (1999).
- Méthot, N. & Basler, K. An absolute requirement for *Cubitus interruptus* in Hedgehog signaling. *Development* **128**, 733–742 (2001).
- Aza-Blanc, P., Ramirez-Weber, F. A., Laget, M. P., Schwartz, C. & Kornberg, T. B. Proteolysis that is inhibited by Hedgehog targets *Cubitus interruptus* protein to the nucleus and converts it to a repressor. *Cell* **89**, 1043–1053 (1997).
- Pearse, R. V. & Tabin, C. J. The molecular ZPA. *J. Exp. Zool.* **282**, 677–690 (1998).
- Yang, Y. *et al.* Relationship between dose, distance and time in *Sonic hedgehog*-mediated regulation of anteroposterior polarity in the chick limb. *Development* **124**, 4393–4404 (1997).
- Lewis, P. M. *et al.* Cholesterol modification of *Sonic hedgehog* is required for long-range signaling activity and effective modulation of signaling by Ptc1. *Cell* **105**, 599–612 (2001).
- Chiang, C. *et al.* Manifestation of the limb prepattern: Limb development in the absence of *Sonic hedgehog* function. *Dev. Biol.* **236**, 421–435 (2001).
- Kraus, P., Fraidenaich, D. & Loomis, C. A. Some distal limb structures develop in mice lacking *Sonic hedgehog* signaling. *Mech. Dev.* **100**, 45–58 (2001).
- Sanz-Ezquerro, J. J. & Tickle, C. "Fingering" the vertebrate limb. *Differentiation* **69**, 91–99 (2001).
- Dai, P. *et al.* *Sonic hedgehog*-induced activation of the *Gli1* promoter is mediated by GLI3. *J. Biol. Chem.* **274**, 8143–8152 (1999).
- Brewster, R., Lee, J. & Ruiz i Altaba, A. Gli/Zic factors pattern the neural plate by defining domains of cell differentiation. *Nature* **393**, 579–583 (1998).
- Brewster, R., Mullor, J. L. & Ruiz i Altaba, A. Gli2 functions in FGF signaling during antero-posterior patterning. *Development* **127**, 4395–4405 (2000).
- Shin, S. H., Kogerman, P., Lindstrom, E., Toftgard, R. & Biesecker, L. G. GLI3 mutations in human disorders mimic *Drosophila* *Cubitus interruptus* protein functions and localization. *Proc. Natl Acad. Sci. USA* **96**, 2880–2884 (1999).
- Sasaki, H., Nishizaki, Y., Hui, C.-c., Nakafuku, M. & Kondoh, H. Regulation of Gli2 and Gli3 activities by an amino-terminal repression domain: implication of Gli2 and Gli3 as primary mediators of Shh signaling. *Development* **126**, 3915–3924 (1999).
- Hui, C. C. & Joyner, A. L. A mouse model of Greig cephalopolysyndactyly syndrome: the *extra-toes*¹ mutation contains an intragenic deletion of the *Gli3* gene. *Nature Genet.* **3**, 241–246 (1993).
- Zúñiga, A. & Zeller, R. *Gli3* (*Xt*) and *formin* (*ld*) participate in the positioning of the polarising region and control of posterior limb-bud identity. *Development* **126**, 13–21 (1999).
- te Welscher, P., Fernandez-Teran, M., Ros, M. A. & Zeller, R. Mutual genetic antagonism involving

GLI3 and dHAND prepatterns the vertebrate limb bud mesenchyme prior to SHH signaling. *Genes Dev.* **16**, 421–426 (2002).

- Litingtung, Y. & Chiang, C. Specification of neuronal cell types in the ventral spinal cord is mediated by antagonistic interaction between Shh and Gli3. *Nature Neurosci.* **3**, 979–985 (2000).
- Caruccio, N. C. *et al.* Constitutive activation of *Sonic hedgehog* signaling in the chicken mutant *talpid*²: Shh-independent outgrowth and polarizing activity. *Dev. Biol.* **212**, 137–149 (1999).
- Martin, G. R. The roles of FGFs in the early development of vertebrate limbs. *Genes Dev.* **12**, 1571–1586 (1998).
- Zúñiga, A., Haramis, A.-P. G., McMahon, A. & Zeller, R. Signal relay by BMP antagonism controls the SHH/FGF4 feedback loop in vertebrate limb buds. *Nature* **401**, 598–602 (1999).
- Sun, X. *et al.* Conditional inactivation of Fgf4 reveals complexity of signalling during limb bud development. *Nature Genet.* **25**, 83–86 (2000).
- Fraidenaich, D., Lang, R. & Basilico, C. Distinct regulatory elements govern Fgf4 expression in the mouse blastocyst, myotomes, and developing limb. *Dev. Biol.* **204**, 197–209 (1998).
- Castilla, E. E. *et al.* Epidemiological analysis of rare polydactylies. *Am. J. Med. Genet.* **65**, 295–303 (1996).
- Qu, S. *et al.* Polydactyly and ectopic ZPA formation in *Alx-4* mutant mice. *Development* **124**, 3999–4008 (1997).
- Radhakrishna, U. *et al.* The phenotypic spectrum of Gli3 morphopathies includes autosomal dominant preaxial polydactyly type IV and postaxial polydactyly type A/B. *Am. J. Hum. Genet.* **65**, 645–655 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank Y.-F. Wang, R. Mernaugh and J. Lancman for technical assistance; R. Harland and B. Vogelstein for reagents; and S. Carroll and members of the Fallon laboratory for critical comments on the manuscript. This work was supported by grants from the National Institutes of Health to C.C. and J.F.F.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.C. (e-mail: chin.chiang@vanderbilt.edu) or J.F.F. (e-mail: jffallon@facstaff.wisc.edu).

T-cell engagement of dendritic cells rapidly rearranges MHC class II transport

Marianne Boes*, Jan Cerny†‡§, Ramiro Massol†‡
Marjolein Op den Brouw*, Tom Kirchhausen†‡, Jianzhu Chen||
& Hidde L. Ploegh*

* Department of Pathology,

† The Center for Blood Research, and

‡ Department of Cell Biology, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA

|| Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, 40 Ames Street, Cambridge, Massachusetts 02139, USA

§ Faculty of Science, Charles University, Albertov 6, Prague, van Černý, Czech Republic

Assembly of major histocompatibility complex (MHC) molecules, which present antigen in the form of short peptides to T lymphocytes, occurs in the endoplasmic reticulum; once assembled, these molecules travel from the endoplasmic reticulum to their final destination. MHC class II molecules follow a route that takes them by means of the endocytic pathway, where they acquire peptide, to the cell surface¹. The transport of MHC class II molecules in 'professional' antigen-presenting cells (APCs) is subject to tight control and responds to inflammatory stimuli such as lipopolysaccharide. To study class II transport in live APCs, we replaced the mouse MHC class II gene with a version that codes for a class II molecule tagged with enhanced green fluorescent protein (EGFP). The resulting mice are immunologically indistinguishable from wild type. In bone-marrow-

derived dendritic cells, we observed class II molecules in late endocytic structures with transport patterns similar to those in Langerhans cells observed *in situ*. We show that tubular endosomes extend intracellularly and polarize towards the interacting T cell, but only when antigen-laden dendritic cells encounter T cells of the appropriate specificity. We propose that such tubulation serves to facilitate the ensuing T-cell response.

To visualize transport of class II MHC molecules in live APCs, we generated mice in which the endogenous class II product is replaced by a fluorescently tagged version, through germline manipulation. We replaced the *I-A^b* gene in J1 (129/Sv; H-2^b) embryonic stem (ES) cells with a construct encoding a protein composed of an intact class II β -chain fused at its carboxy terminus with an EGFP moiety (Fig. 1a). One homologous integrant embryonic stem cell clone was obtained and used to derive germline knock-in mice, identified by fluorescence-activated cell sorting (FACS) screening of I-A^b EGFP-positive B cells in peripheral blood (data not shown). Knock-in mice, but not wild-type animals, showed the presence of an anti-EGFP-reactive polypeptide of the expected size, 56K (relative molecular mass (M_r) of 56,000; Fig. 1b). No free EGFP (27K) was detected. Thus, EGFP visualized by microscopy represents MHC class II molecules tagged with EGFP (MHC II-EGFP).

Maturation of MHC class II molecules, monitored by pulse-chase analysis and digestion with endoglycosidase H, occurs normally in MHC II-EGFP mice. The rate of acquisition of endoglycosidase H

resistance is indistinguishable for MHC class II molecules and the EGFP-tagged class II molecules (data not shown). The diagnostic presence of SDS-resistant dimers—a surrogate marker for peptide occupancy—indicates that peptide loading proceeds with similar kinetics in heterozygous and in homozygous MHC II-EGFP mice (Fig. 1c).

The expression of MHC products is a prerequisite for development of an appropriately restricted T-cell repertoire^{2,3}. Flow cytometry of thymus and lymph node showed normal production of CD4⁺ and CD8⁺ T cells in MHC II-EGFP mice (Fig. 1d). We found no significant differences in serum immunoglobulin levels between the wild-type and knock-in mice (Fig. 1e). The MHC II-EGFP fusion protein therefore allows APCs to stimulate the CD4 T-cell populations required for immunoglobulin class switching.

We generated dendritic cells from MHC II-EGFP mice using bone marrow cultures supplemented with granulocyte/macrophage-colony stimulating factor (GM-CSF) and interleukin (IL)-4. At day 5, yields of dendritic cells and marker profiles (CD11b, CD11c, CD80, CD86, MHC class II) were indistinguishable from wild-type littermates (data not shown). Such dendritic cells expressed low to intermediate levels of CD86, and MHC II-EGFP was contained almost entirely intracellularly. There is a strict requirement for the activation of dendritic cells, notably by microbial products such as lipopolysaccharide (LPS), in order to achieve strong surface expression of MHC class II molecules²². Indeed, treatment of

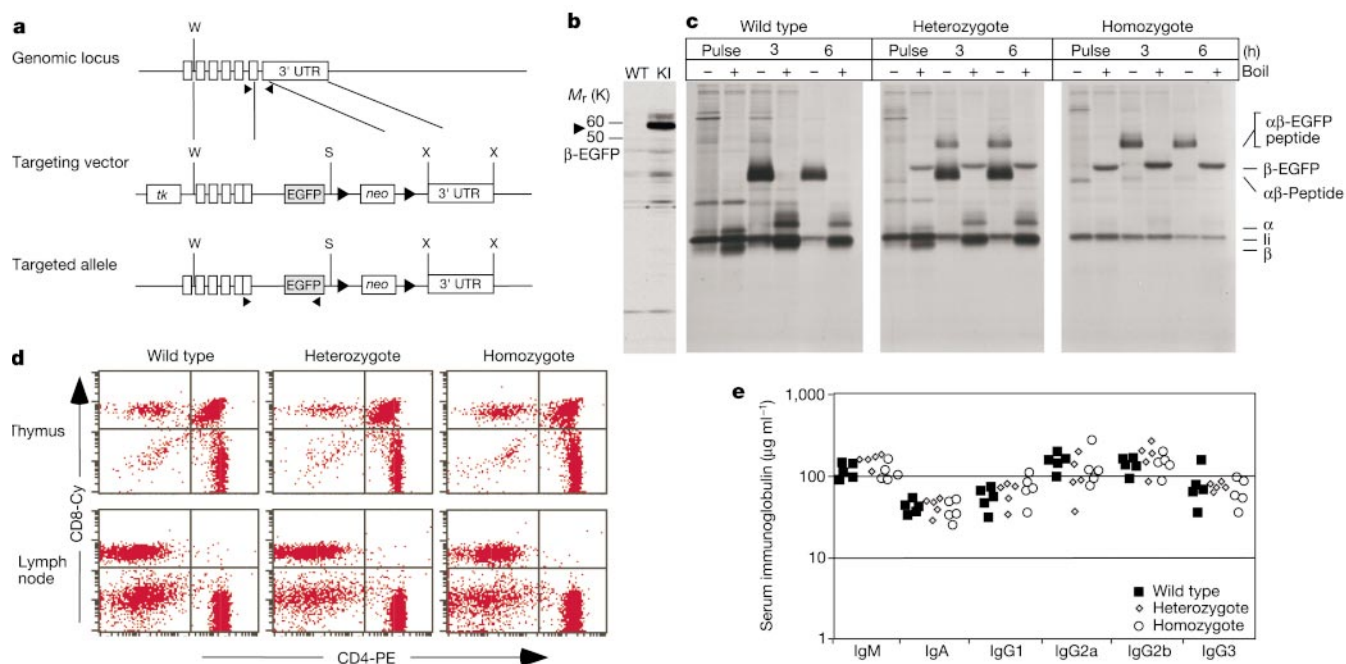


Figure 1 Generation and characterization of MHC II-EGFP knock-in mice. **a**, I-A^b-EGFP knock-in mice (H-2^b background) were generated by replacing the β -chain gene with β -EGFP via homologous recombination in ES cells. The genomic sequence encoding exons 2–6 were used as the 5' homologous region. The 3' untranslated region was used as the 3' homologous region, obtained by PCR from genomic DNA. The targeting vector consisted of a floxed (triangles) phosphoglycerate kinase promoter-driven neomycin (*neo*) resistance gene, flanked by a 4.8-kb *Swat/SatI* (S) homologous fragment upstream and by a 1.6-kb *XbaI/XbaI* (X) homologous fragment downstream. As a marker for counter selection, the thymidine kinase (*tk*) gene was placed upstream from the 5' homologous region. Sequences encoding exons 5 and 6 were ligated together using a linker sequence (5'-TCGATTGTCGACATTTAAATCTCGAGGCCCTCTCCAGCAGGACTCCTGCAG-3'). **b**, Whole spleen cell lysates were separated by 12.5% SDS-PAGE and transferred to PVDF membranes. Membranes were blotted using a polyclonal anti-EGFP antibody generated against bacterially expressed EGFP; arrowhead indicates position of fusion

protein. WT, wild type; KI, knock-in. **c**, Splenocytes were pulse-labelled for 45 min with 0.5 mCi ml⁻¹ [³⁵S]methionine/cysteine and chased in complete DMEM for 3 and 6 h. Class II molecules were recovered by immunoprecipitation with the N22 antibody. Immunoprecipitates were re-suspended in sample buffer containing 5% 2-ME, either boiled or kept on ice for 5 min, and analysed by 12.5% SDS-PAGE. Ii, invariant chain. **d**, T-cell populations in knock-in mice analysed by flow cytometry. Single-cell suspensions prepared from thymus and lymph node were stained with conjugated antibodies and analysed by FAC-Scan (Becton Dickinson). Percentages of CD4 T cells are: thymus, 27.4%, 27.5% and 38.8%; lymph node, 51.6%, 52.9% and 58.7% for wild-type, heterozygous and homozygous knock-in animals, respectively. Percentages of CD8 T cells are: thymus, 7.6%, 5.5% and 8.7%; lymph node, 23.6%, 28.4% and 15.8% for wild-type, heterozygous and homozygous knock-in animals, respectively. **e**, Serum immunoglobulin levels determined by ELISA from 6-week-old littermates. Each symbol represents a serum sample from one mouse.

dendritic cells with LPS led to a redistribution of MHC II-EGFP molecules to the plasma membrane (Fig. 2d, right panel).

In dendritic cells we observed extensive co-localization of MHC II-EGFP with internalized DiI-labelled low-density lipoprotein (LDL), a protein targeted ultimately for lysosomal destruction (See Supplementary Information 1 and ref. 4). In contrast, there was little apparent co-localization of Alexa Fluor 594-labelled transferrin with MHC II-EGFP (Fig. 2a). Thus, the class II-positive structures are not readily accessible to tracers that mark early and recycling endocytic compartments. We then exposed dendritic cells to *Salmonella typhimurium* expressing the red fluorescent protein DsRed1. After 3 h, internalized bacteria did not show appreciable co-localization with MHC II-EGFP (Fig. 2b), but after 5 h each intracellular bacterium resided in a compartment positive for MHC II-EGFP (Fig. 2c). Combined, these data establish that MHC II-EGFP molecules reside primarily in late endocytic compartments.

Intracellular movement of MHC II-EGFP-positive compartments was quite dynamic (see also ref. 5), as changes in location can be observed in confocal images captured only seconds apart. EGFP-positive compartments are vesicular and tubular structures, with tubules showing the most dynamic movement (Fig. 2d). These tubules are up to approximately $5 \mu\text{m}$ in length, move at speeds up to $2 \mu\text{m s}^{-1}$ (Supplementary Information 2d) and may be similar to those reported in D1 cells exposed for several hours to LPS⁶. This transport of class II-positive compartments is strictly microtubule-dependent. Within 8 min of application of the microtubule-disrupting agent nocodazole, class II-EGFP-positive tubules are no longer visible in the more distal projections of the cells. After 40 min of nocodazole treatment, all MHC class II molecules had collapsed to a location close to the nucleus (Supplementary Information 2).

Langerhans cells are regarded as prototypic immature dendritic cells localized in the epidermis. In fixed tissue sections, Langerhans cells are strongly positive for MHC class II molecules (ref. 7). We examined the distribution of MHC II-EGFP in fresh epidermal sheets prepared from mouse ears. A three-dimensional reconstruction of an isotropic set of 50 confocal images shows the distribution of Langerhans cells within the epidermis (Fig. 3a). We found that MHC II-EGFP is localized primarily to the central portion of the cell, with well-demarcated vesicles present in the more remote extensions, giving the appearance of beads on a string. Intracellular movement of class II-positive compartments can be observed *in situ* even in these preparations (Fig. 3b).

We next studied the fate of class II-containing compartments in dendritic cells on contact with T cells. We pulsed dendritic cells with either hen egg lysozyme (HEL) or ovalbumin (Ova) overnight and removed excess antigen by washing. Under these circumstances, display at the cell surface of peptide-loaded MHC molecules will occur. We then added the HEL-specific T hybridoma 46.13 (HEL, residues 46–61; I-A^b restricted)⁸, or the Ova-specific T hybridoma B097.1 (Ova, residues 323–339; I-A^b restricted)⁹. Randomly selected dendritic cell–T-cell conjugates were then analysed immediately by time-lapse confocal microscopy (Fig. 4a, b). Contacts between T cells and dendritic cells resulted in immediate (within minutes) and pronounced tubulation of class II-positive compartments, such that a tubule (usually one or two per cell) would point directly to the area of contact between a dendritic cell and a T cell. These tubules reached extraordinary lengths, at times exceeding $50 \mu\text{m}$, and emanated from the large accumulation of MHC II-EGFP around the microtubule-organizing centre. Both the kinetics of tubulation and the length of tubules formed are quite different from those seen in the D1 cell line exposed to LPS⁶. Where several T cells interacted with a single dendritic cell, tubules extended to each of the interacting T cells. Tubules of much reduced length and of no preferred orientation were observed when dendritic cells had not encountered antigen, yet were exposed to antigen-specific T cells in the same manner (data not shown; see

also Fig. 5). Samples examined at later times (2 and 6 h after addition of antigen-specific T cells) still contained such extended tubular class II-positive compartments in 55–75% of dendritic cells exposed to antigen, but in fewer than 5% of dendritic cells loaded with the wrong antigen or with no antigen at all. Three-dimensional reconstructions performed on fixed dendritic cell–T-cell conjugates showed the presence of MHC II-EGFP-positive tubules laid down on a lattice of microtubules (Fig. 4c), which was again disrupted by nocodazole (data not shown).

Primary T cells obtained from the lymph nodes of OTII T-cell

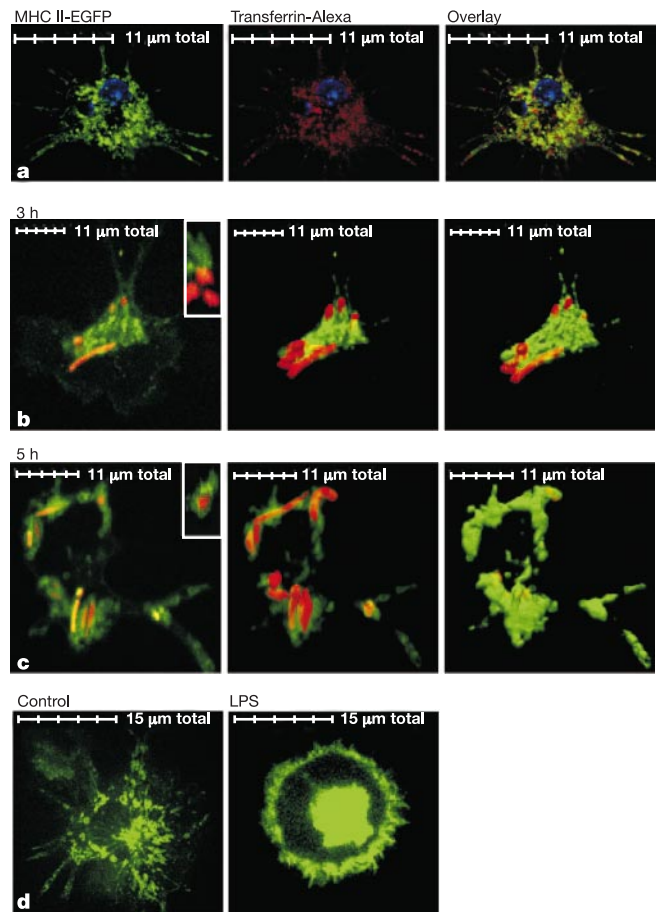


Figure 2 MHC II-EGFP molecules in dendritic cells reside in late endosomes. **a**, Dendritic cells (day 5 of culture) were allowed to endocytose transferrin labelled with Alexa Fluor 594 for 30 min, and were then immediately visualized in a single optical section by time-lapse confocal microscopy. Nuclei (blue) are visualized by incubation with Hoechst 33258 (15 min at 37°C). **b**, Dendritic cells (day 5) were allowed to endocytose live *S. typhimurium* expressing DsRed1 for 1 h. Unbound bacteria were washed away, and dendritic cells were analysed by confocal microscopy after a total of 3 h (left). Inset, magnification of a cross-section taken perpendicular to the confocal image. *S. typhimurium* is close to, but not enveloped by, a MHC II-EGFP⁺ compartment (middle). After 3 h, bacteria are taken up by dendritic cells, but do not reside in MHC II-EGFP⁺ compartments (right). **c**, After 5 h of exposure, most *S. typhimurium* are enveloped within MHC II-EGFP⁺ compartments. Dendritic cells were analysed by confocal microscopy (left). Inset, magnification of a cross-section taken perpendicular to the confocal plane shown in the larger image. *S. typhimurium* is surrounded by a MHC II-EGFP⁺ compartment. By rendering the green fluorescence either transparent (middle) or opaque (right), the complete engulfment of *S. typhimurium* at 5 h is apparent. At 5 h, degradation of bacteria commences, as in some vesicles the DsRed1-labelled bacteria appear as round rather than rod-like structures. **d**, Day 4 dendritic cells were either left untreated (left) or were treated with $5 \mu\text{g ml}^{-1}$ LPS overnight (right) and visualized the next day by confocal microscopy in a single optical section. LPS activation induced morphological changes and increased display of class II molecules on the cell membrane (right).

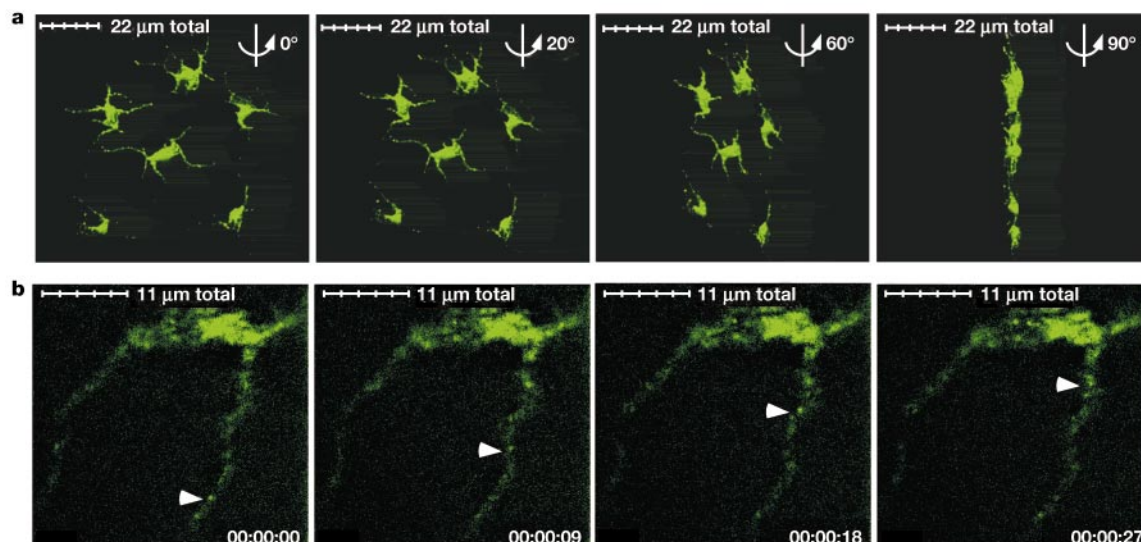


Figure 3 Transport of MHC II-EGFP-containing structures in Langerhans cells *in situ*. **a**, Langerhans cells on the epidermal sheet were visualized by three-dimensional volume renditions of 50 confocal microscopy images. **b**, Examination of a single Langerhans cell

in a whole mount by time-lapse microscopy in a single optical section. Vesicular movement is apparent, as evident from the tracking of a single vesicle (arrowhead).

antigen receptor (I-A^b restricted, Ova-specific) transgenic mice (Fig. 5) evoked similar extensive tubulation (again, one to two extended tubules per cell) when confronted with Ova-loaded, but not control, dendritic cells (Fig. 5c). A quantification of the length, orientation and particularly the frequency of the polarized tubules aimed at the T cell (Fig. 5b, category 1 tubules) illustrates the

marked response elicited by naive OTII T cells when cognate antigen is present. OTII T cells exposed to antigen-laden dendritic cells upregulated CD69, but not when Ova was withheld (Fig. 5a). Activation of naive OTII T cells thus correlates with the antigen-dependent occurrence of the tubulation induced by them.

The availability of mice that express EGFP-tagged MHC class II

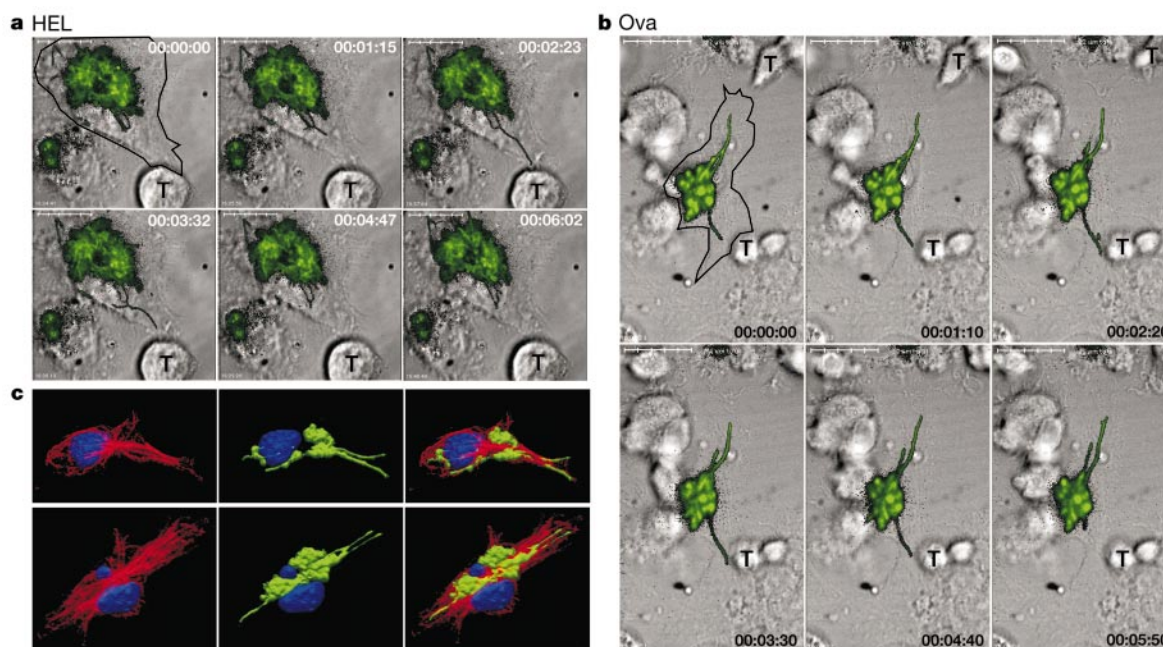


Figure 4 Microtubule-driven directional transport of endosomal compartments on T-cell contact. **a**, Day 4 dendritic cells cultured on coverslips at 1×10^6 per well (six-well plate) and incubated with 300 μ M HEL (Sigma) overnight to induce display of HEL-peptide on class II molecules on the cell membrane. T hybridoma cells specific for HEL(46–61) peptide were brought into contact with dendritic cells by a 30-s centrifugation step (see Methods). Time-lapse confocal microscopy in a single optical section was performed immediately after imposition of contact, and at 2 and 6 h thereafter (not shown), yielding most MHC II-EGFP tubuli after 2 and 6 h. Scale bar, 13 μ m. **b**, As in **a**, except that dendritic cells were incubated with 300 μ M Ova and T hybridoma cells specific to

OVA(323–339) were used. Most MHC II-EGFP tubuli were revealed after 2 and 6 h. A T cell in the upper right quadrant is moving out of the field of view between 2.20 and 3.30 min, whereas a T cell in the lower right corner is stationary for the time observed. Scale bar, 22 μ m. **c**, T hybridoma cells specific for HEL(46–61) peptide were added to HEL-loaded dendritic cells and were cultured for an additional 2 h at 37 °C to induce tubulation of MHC II-EGFP⁺ compartments. Dendritic cell/T-cell clusters were fixed in 3% paraformaldehyde, and microtubuli were visualized by staining for α -tubulin (red, labelled with Cy3, nuclei are stained blue; see Methods).

molecules has allowed us to visualize the transport of class II MHC molecules in live cells. The use of a knock-in approach ensures balanced levels of expression of the MHC II-EGFP reporter in the appropriate cell types. Moreover, it allows an *in vivo* validation of the functionality of the EGFP-tagged class II molecules by examining mice homozygous for the knock-in mutation. This is important, because even non-professional APCs such as fibroblasts can be rendered capable of antigen presentation and T-cell stimulation by simple transfection with class II complementary DNAs¹⁰, yet few

would argue that such cells should be considered representative of professional APCs.

Having established that MHC II-EGFP mice are phenotypically normal, the knock-in mice were used as a source of bone marrow from which dendritic cells could be cultured by providing GM-CSF and IL-4 growth factors. On contacting an APC, T cells form supramolecular activation clusters, or SMACs, also referred to as the immunological synapse^{11,12}. Our results indicate marked changes in subcellular organization in the APC also: tubular endosomes that contain abundant class II molecules are aimed directly at the interacting T cell in a strictly antigen-dependent manner (Figs 4 and 5). Such rearrangement may maximize delivery of peptide-loaded class II molecules to the T cell. Models for T-cell activation prominently include a requirement for multiple sequential interactions for the T cell to surpass the threshold indicative of a productive interaction with the APC^{13–15}. The ability of a dendritic cell to direct and sustain delivery of class II peptide complexes to site(s) at which the interactions with T cells occur would assist in generating a productive signal. Prolonged interaction (hours) between dendritic cells and T cells may be required to obtain full activation¹⁶. How can the initial dendritic cell–T-cell contact be used to maximize delivery of MHC peptide complexes to the responding T cell? Our observations suggest that directed transport of class II molecules from intracellular stores to the T-cell contact site(s) may provide the solution.

There may well be a link between the ligation of class II molecules at the surface, caused by contact with an antigen-specific T cell, and the ensuing polarization towards the contact interface. Specifically, the activation of protein kinase C occurs on ligation of surface class II molecules¹⁷. For such signalling to be possible, some small number of the appropriate peptide–MHC complexes would have to be present. Further recruitment of class II molecules from intracellular stores would follow by tubulation on T-cell contact. If, as we suspect, the rearrangement of the class II transport pathway increases a T cell's probability of receiving appropriate stimulation, then microbes could target this process as a means of immune evasion. Many bacteria have co-opted the host cytoskeleton as essential for their survival or contributing to virulence. It would not seem far-fetched, then, to suggest interference with tubulation as a possible escape mechanism. □

Methods

Targeting vector and mutant mice

To produce a mutant mouse strain carrying an EGFP label on all MHC β -chains, we made use of an ES cell line, J1, which was derived from a 129/Sv mouse. This strain is b-haplotype at the H-2 locus and therefore carries a mutation in the *E α* gene promoter that precludes synthesis of E α protein, and therefore prevents expression of cell surface I-E complex¹⁸. The starting material was a 6.3-kilobase (kb) *EcoRI* fragment from an A β genomic clone carried in the plasmid pcEXV-n (a gift from D. Mathis).

Western blot

Whole-cell lysates were prepared from 1×10^7 splenocytes by lysis in NP-40 lysis buffer pH 7.4 (50 mM Tris-HCl, 150 mM NaCl, 0.5% NP-40) supplemented with 1% SDS for 30 min on ice. Proteins were separated on SDS–polyacrylamide gel electrophoresis (PAGE) and transferred to polyvinylidene difluoride (PVDF) membranes (NEN Life Science Products). Membranes were blocked with 10% milk in PBS/0.1% Tween, and blotted with a rabbit polyclonal anti-EGFP antibody that was generated by our laboratory.

Pulse-chase analysis and immunoprecipitations

Pulse-chase experiments and immunoprecipitations were performed as described¹⁹. Briefly, splenocytes were pulsed for 45 min with 0.5 mCi ml^{-1} [³⁵S]methionine/cysteine (Du Pont New England Nuclear) and chased in complete DMEM for 3 and 6 h. After each chase point, cells were lysed in NP-40 lysis buffer at pH 7.4 supplemented with 1% SDS. Class II molecules were recovered by immunoprecipitation with the N22 antibody; a reagent that captures properly folded and assembled class II molecules (N22 was a gift of R. M. Steinman). Immunoprecipitations were carried out with *Staphylococcus aureus*, were washed four times with washing buffer (0.5% NP-40, 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 5 mM EDTA), resuspended in sample buffer containing 5% 2-ME, either boiled or kept on ice for 5 min, and analysed by SDS–PAGE.

Flow cytometry analyses

CD4 and CD8 T-cell development in the knock-in mice was analysed by flow cytometry.

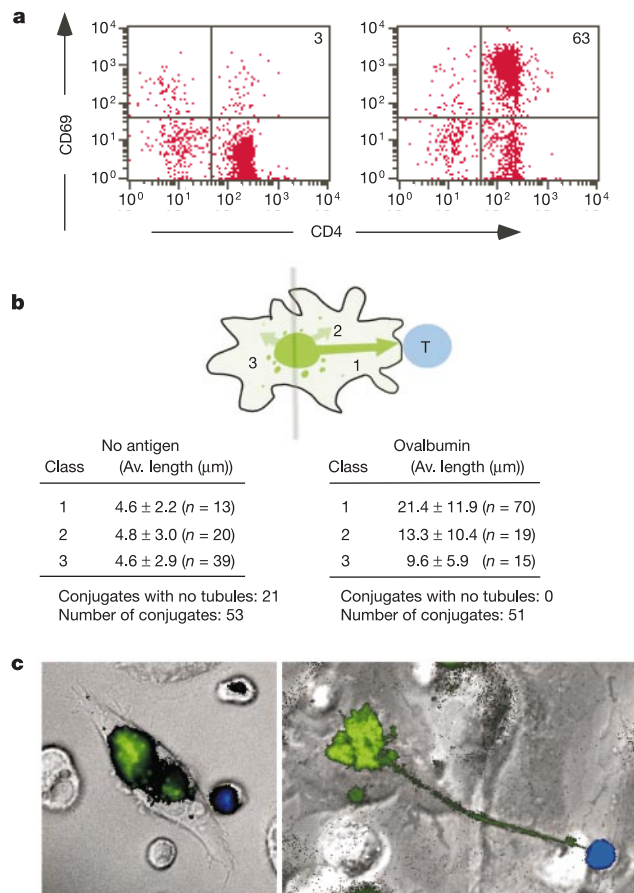


Figure 5 Primary T cells induce tubulation in dendritic cells in an antigen-dependent manner. **a**, Analysis of T-cell activation by expression of the CD69 marker. OTII T cells (lymph node) were exposed to antigen-pulsed dendritic cells (day 5) for 18 h and analysed by cytofluorimetry for CD4 and CD69 markers. Numbers in the upper-right quadrant indicate cell percentage in the quadrant. **b**, Quantitative analysis of tubulation induced by OTII cells. Naive OTII T cells were labelled with Hoechst 33258 and incubated with dendritic cells (day 5) in the absence or presence of Ova (200 μM, 18 h). OTII T cells were added to the coverslips on which dendritic cells were grown and conjugate formation was initiated by a brief centrifugation. Images were captured in epifluorescence mode. Tubules that extended directly towards the interacting T cell were classified as '1', those that included a vector towards the interacting T cell, but not pointing directly at the T cell, were classified as '2', and tubules pointing away from the interacting T cell were classified as '3' (top). For some conjugates, more than one tubule extending towards the T cell was observed. In the absence of antigen, a conjugate was defined as a dendritic cell touched by a T cell. Numbers of conjugates, numbers of tubules, mean values for tubule length and s.d. are given. *t*-test analysis of the tubules classified as Ova-treated compared with untreated dendritic cells yields the *P*-values: 1, *P* < 0.0001; 2, *P* < 0.0001; 3, *P* = 0.0002. **c**, Tubulation in the absence and presence of Ova. Superposition of the bright-field image, green fluorescence (dendritic cells) and blue fluorescence (T cell) is shown. The left panel shows tubulation in the absence of Ova; tubulation in the presence of Ova is shown in the right panel. The length of the MHC class II⁺ tubule is 41 μm. Images were taken 3 h after addition of T cells.

Phycoerythrin- and cychrome-conjugated monoclonal antibodies specific for CD4 and CD8 were from PharMingen. Single-cell suspensions were prepared from thymus and lymph node. Ten thousand live cells were collected for each sample.

Enzyme-linked immunosorbent assay (ELISA)

We measured serum immunoglobulin levels by ELISA (Southern Biotechnology Associates). Plates were coated with goat anti-mouse immunoglobulin- μ , γ and α and developed with horseradish peroxidase-conjugated goat antibodies specific for each mouse immunoglobulin isotype. The concentrations were calculated using the linear ranges of the dilution and purified mouse immunoglobulin isotypes as standards.

Cell preparation and culture

Dendritic cells were generated as described²⁰. Briefly, cells were flushed from the bone marrow cavity with PBS/2.5% FCS. Cells were plated at 1×10^6 cells per well in 2 ml DMEM, 25 mM HEPES and 10% FCS without phenol red, and 10 ng ml⁻¹ GM-CSF and 1 ng ml⁻¹ IL-4. Cells were cultured on 25-mm circular coverslips, with changes of medium every second day; non-adherent cells were removed by gentle washing. For activation, LPS (5 μ g ml⁻¹) was added for the last 20 h of culture.

Epidermal sheet preparation

Ear snips were obtained from MHC II-EGFP mice, and hairs were removed using an epilating agent (Nair, Carter Products). Epidermal sheets were generated by separation of ear snips into dorsal and ventral halves, with the use of forceps.

Imaging of MHC II-EGFP molecules in dendritic cells

We used 5-day-old bone marrow dendritic cells for real-time imaging. For dissection of early and late endocytic compartments, dendritic cells were incubated with 25 μ g ml⁻¹ transferrin Alexa Fluor 594 (Molecular Probes) or DiI-LDL (10 μ g ml⁻¹; provided by M. Krieger) for 30 min at 37 °C in DMEM without serum and washed three times. Similarly, dendritic cells were incubated with DsRed1-labelled (Clontech) *S. typhimurium* (provided by M. Starnbach) in normal medium without penicillin/streptomycin for 1 h, and analysed after a total of 3–5 h. Images were captured with an inverted Zeiss 200M microscope equipped with a $\times 63$ objective (1.4NA PanApo) and a spinning-wheel confocal head (Perkin-Elmer), and supplied with a 37 °C open perfusion temperature-controlled chamber (20/20 Technology). We used Slidebook (Intelligent Imaging Innovation Inc) for image acquisition and data processing. Volocity (Improvision Inc) was used for creating a three-dimensional rendition of the volume.

Antigen presentation in real time

Dendritic cells at a concentration of 1×10^6 per coverslip were incubated with 300 μ M HEL or Ova (Sigma) to induce display of antigen-derived peptides on class II molecules. Dendritic cells were then washed three times with culture medium. T cells (0.5×10^6) were labelled with the nuclear dye Hoechst 33258 (Molecular Probes), washed and then added to each coverslip containing dendritic cells. The HEL-specific T-cell hybridoma used was H46.13, which is specific for HEL(46–61) (provided by C. Alfonso); the Ova-specific T-cell hybridoma was B097.1, which is specific for Ova(323–339) (provided by P. Marrack). Freshly isolated Ova-specific OTII T cells were obtained from lymph nodes from OTII transgenic mice²¹. T cells were brought into contact with dendritic cells by a quick centrifugation step at 600 g. at room temperature. Confocal microscopy analysis was performed immediately after 2 and 6 h.

Immunostaining of fixed dendritic cells

Five-day-old dendritic cells on coverslips were fixed in 3% paraformaldehyde in PBS, 1 mM Mg²⁺ and 0.1 mM Ca²⁺ (PBS-MC) for 10 min at 37 °C. Dendritic cells were washed extensively with PBS-MC and were permeabilized for 10 min with 0.5% saponin in PBS-MC at room temperature. Dendritic cells were washed and any excess of reactive groups of paraformaldehyde were quenched with 50 mM NH₄Cl in PBS-MC for 15 min at room temperature. Anti- α tubulin (1:100, Oxford Biotechnology) staining was done in PBS-MC overnight at 4 °C. We used a rat anti-mouse Cy3-labelled secondary antibody for detection (Jackson ImmunoResearch).

Received 4 April; accepted 3 July 2002; doi:10.1038/nature01004.

- Villadangos, J. A. *et al.* Proteases involved in MHC class II antigen presentation. *Immunol. Rev.* **172**, 109–120 (1999).
- Cosgrove, D. *et al.* Mice lacking MHC class II molecules. *Cell* **66**, 1051–1066 (1991).
- Grusby, M. J., Johnson, R. S., Papaioannou, V. E. & Glimcher, L. H. Depletion of CD4 + T cells in major histocompatibility complex class II-deficient mice. *Science* **253**, 1417–1420 (1991).
- Brown, M. S. & Goldstein, J. L. A receptor-mediated pathway for cholesterol homeostasis. *Science* **232**, 34–47 (1986).
- Wubbolts, R. *et al.* Direct vesicular transport of MHC class II molecules from lysosomal structures to the cell surface. *J. Cell Biol.* **135**, 611–622 (1996).
- Kleijmeer, M. *et al.* Reorganization of multivesicular bodies regulates MHC class II antigen presentation by dendritic cells. *J. Cell Biol.* **155**, 53–63 (2001).
- Rowden, G. The Langerhans cell. *Crit. Rev. Immunol.* **3**, 95–180 (1981).
- Maric, M. *et al.* Defective antigen processing in GILT-free mice. *Science* **294**, 1361–1365 (2001).
- Hugo, P., Kappler, J. W., Godfrey, D. I. & Marrack, P. C. A cell line that can induce thymocyte positive selection. *Nature* **360**, 679–682 (1992).
- Malissen, B., Steinmetz, M., McMillan, M., Pierres, M. & Hood, L. Expression of I-Ak class II genes in mouse L cells after DNA-mediated gene transfer. *Nature* **305**, 440–443 (1983).
- Monks, C. R., Freiberg, B. A., Kupfer, H., Sciaky, N. & Kupfer, A. Three-dimensional segregation of supramolecular activation clusters in T cells. *Nature* **395**, 82–86 (1998).
- Grakoui, A. *et al.* The immunological synapse: a molecular machine controlling T cell activation. *Science* **285**, 221–227 (1999).

- Valitutti, S., Muller, S., Cella, M., Padovan, E. & Lanzavecchia, A. Serial triggering of many T-cell receptors by a few peptide-MHC complexes. *Nature* **375**, 148–151 (1995).
- Underhill, D. M., Bassetti, M., Rudensky, A. & Aderem, A. Dynamic interactions of macrophages with T cells during antigen presentation. *J. Exp. Med.* **190**, 1909–1914 (1999).
- Gunzer, M. *et al.* Antigen presentation in extracellular matrix: interactions of T cells with dendritic cells are dynamic, short lived, and sequential. *Immunity* **13**, 323–332 (2000).
- Lanzavecchia, A. & Sallusto, F. Regulation of T cell immunity by dendritic cells. *Cell* **106**, 263–266 (2001).
- Scholl, P. R. & Geha, R. S. MHC class II signalling in B-cell activation. *Immunol. Today* **15**, 418–422 (1994).
- Mathis, D. J., Benoist, C., Williams, V. E., Kanter, M. & McDevitt, H. O. Several mechanisms can account for defective E α gene expression in different mouse haplotypes. *Proc. Natl Acad. Sci. USA* **80**, 273–277 (1983).
- Villadangos, J. A., Riese, R. J., Peters, C., Chapman, H. A. & Ploegh, H. L. Degradation of mouse invariant chain: roles of cathepsins S and D and the influence of major histocompatibility complex polymorphism. *J. Exp. Med.* **186**, 549–560 (1997).
- Villadangos, J. A. *et al.* MHC class II expression is regulated in dendritic cells independently of invariant chain degradation. *Immunity* **14**, 739–749 (2001).
- Barnden, M. J., Allison, J., Heath, W. R. & Carbone, F. R. Defective TCR expression in transgenic mice constructed using cDNA-based alpha- and beta-chain genes under the control of heterologous regulatory elements. *Immunol. Cell Biol.* **76**, 34–40 (1998).
- Turley, S. J. *et al.* Transport of peptide-MHC class II complexes in developing dendritic cells. *Science* **288**, 522–527 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank M. Boxem for help in designing the targeting construct and T. Schmidt for blastocyst microinjections. We thank A. W. M. van der Velden for providing DsRed1-labelled *S. typhimurium*, and we acknowledge discussions with members of the Ploegh laboratory.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.L.P. (e-mail: ploegh@hms.harvard.edu).

Dendritic cell maturation triggers retrograde MHC class II transport from lysosomes to the plasma membrane

Amy Chow, Derek Toomre, Wendy Garrett & Ira Mellman

Department of Cell Biology and Section of Immunobiology, Ludwig Institute for Cancer Research, Yale University School of Medicine, 333 Cedar Street, PO Box 208002, New Haven, Connecticut 06520-8002, USA

Central to the initiation of immune responses is recognition of peptide antigen by T lymphocytes. The cell biology of dendritic cells makes them ideally suited for the essential process of antigen presentation¹. Their life cycle includes several stages characterized by distinct functions and mechanisms of regulation². Immature dendritic cells synthesize large amounts of major histocompatibility complex class II molecules (MHC II), but the $\alpha\beta$ -dimers are targeted to late endosomes and lysosomes (often referred to as MHC class II compartments) where they reside unproductively with internalized antigens. After exposure to microbial products or inflammatory mediators, endocytosis is downregulated, the expression of co-stimulatory molecules is enhanced, and newly formed immunogenic MHC II-peptide complexes are transported to the cell surface^{3–10}. That these MHC II molecules reach the surface is surprising, as the lysosomes comprise the terminal degradative compartment of the

endocytic pathway from which exogenous components generally cannot be recovered intact¹¹. Here we have visualized this pathway in live dendritic cells by video microscopy, using cells expressing MHC II tagged with green fluorescent protein (GFP). We show that on stimulation, dendritic cells generate tubules from lysosomal compartments that go on to fuse directly with the plasma membrane.

Immunofluorescence microscopy of fixed cells shows that MHC class II-peptide complexes form in late endosomes and lysosomes on maturation of dendritic cells, and subsequently in a non-lysosomal compartment of class II vesicles and on the plasma membrane⁸. Electron microscope analysis of a dendritic-cell-like cell line further suggests activation correlated with the marked tubulation of lysosomal membranes¹². The experiments reported here were undertaken to establish the kinetic relationship and functional significance of these events. GFP-tagged MHC II molecules were expressed in live bone-marrow-derived dendritic cells using a recombinant retrovirus. In the mouse, MHC II molecules are encoded by two loci, *I-A* and *I-E*. C57BL/6 mice have a

mutation that abolishes production of the *I-E* α chain¹³. The *I-E* β chain is produced normally, but in the absence of endogenous *I-E* α it is retained in the endoplasmic reticulum (ER). We exploited this knockout and expressed a complementing *I-E* α chain containing a cytoplasmic enhanced GFP tag (EGFP).

Dendritic cell cultures were infected on day 2 while the cells were still proliferating progenitors; expression was assayed on day 4 after differentiation into immature dendritic cells. Viral transduction was efficient, with about 50% of the CD11c-positive dendritic cells (32% of all cells) being GFP positive (Fig. 1a, left). To determine whether the *I-E* α -GFP chain formed a dimer with endogenous *I-E* β , we stained the infected cell populations with an anti-MHC II antibody (14.4.4S) that recognizes only the paired *I-E* $\alpha\beta$ complex¹⁴. Properly assembled *I-E* $\alpha\beta$ dimers were found both by fluorescence-activated cell sorting (FACS) (Fig. 1a, right) and by immunofluorescence. As shown in Fig. 1b, the GFP signal co-localized with 14.4.4S predominantly in intracellular structures (see below). In contrast, infected MHC II-negative cells exhibited the unpaired GFP-tagged α -chain in an ER-like pattern (Fig. 1b). Moreover, about 50% of the

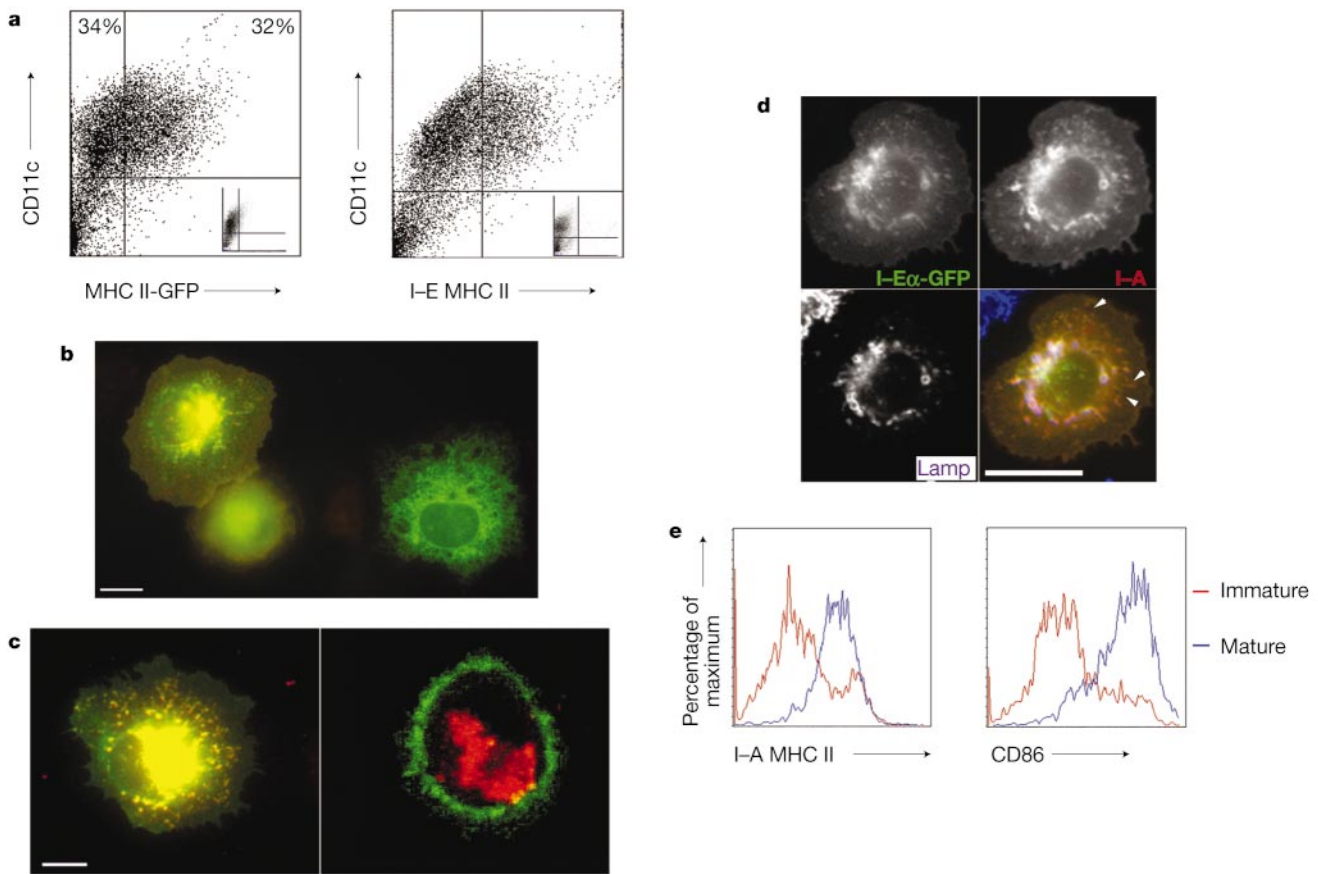


Figure 1 MHC II-GFP expressed by dendritic cells exhibit properties characteristic of endogenous MHC II. **a**, CD11c⁺ dendritic cells express MHC II-GFP and localize I-E $\alpha\beta$ to the cell surface. Transduced dendritic cells were stained with cyochrome-conjugated anti-CD11c alone (left) or together with phycoerythrin (PE)-conjugated 14.4.4S (which recognizes the I-E $\alpha\beta$ complex) (right), and analysed by flow cytometry. Total population percentages are indicated for upper quadrants (left). Insets show the same FACS profiles of untransduced dendritic cells. **b**, 14.4.4S-reactive I-E $\alpha\beta$ localizes intracellularly with MHC II-GFP. Optically merged fixed cell images show I-E $\alpha\beta$ complexes depicted in red and MHC II-GFP in green. I-E $\alpha\beta$ complexes co-localized with MHC II-GFP (yellow) in a dendritic cell, whereas a reticular ER-like pattern of MHC II-GFP was seen in a transduced cell negative for endogenous MHC II (not a dendritic cell). Scale bar, 10 μ m. **c**, MHC II-GFP exhibited proper cellular localization in immature and mature dendritic cells. The left panel shows that MHC II-GFP accumulated predominantly in lysosomes in

immature dendritic cells. The optically merged image shows an immature dendritic cell with Lamp-2 depicted in red and MHC II-GFP in green. Lamp-2 localizes together with MHC II-GFP (yellow). The right panel shows that live mature dendritic cells express high levels of MHC II-GFP on the cell surface with a clustered lysosomal distribution. The optically merged confocal image shows a mature dendritic cell with Texas red-dextran-loaded lysosomes in red and MHC II-GFP in green. Scale bar, 10 μ m. **d**, MHC II-GFP localizes together with endogenous I-A MHC II in class II vesicles after stimulation. White arrowheads in the optically merged three-colour image indicate non-lysosomal compartments in which MHC II-GFP (green) and I-A (red) localize together. Scale bar, 10 μ m. **e**, MHC II-GFP-expressing dendritic cells exhibit normal maturation markers. FACS histogram overlays show transduced dendritic cells labelled with PE-conjugated anti-I-A^b or PE-conjugated anti-CD86. Acquisitions were gated on MHC II-GFP⁺ populations. Red line, immature dendritic cells; blue line, mature dendritic cells.

GFP-tagged MHC II molecules formed a dimer that was stable to SDS at room temperature (detected by western blot with an anti-GFP antibody), suggesting that the MHC II-GFP molecule was peptide loaded (not shown).

Although the I-E α -EGFP chain formed $\alpha\beta$ -dimers, it was essential to determine whether the tagged MHC II molecules exhibited the same intracellular distribution as their endogenous counterparts. As expected, in immature dendritic cells MHC II-GFP co-localized extensively with Lamp-2 (also known as Igp-B), a marker of the lysosomal membrane (Fig. 1c, left). Little MHC II-GFP was found at the cell surface. A markedly different distribution was found after treatment with lipopolysaccharide (LPS). In live mature dendritic cells, lysosomes were depleted of MHC II-GFP, most of which was now at the plasma membrane (Fig. 1c, right), as found previously in fixed cells^{5,8}. Live confocal imaging revealed that surface MHC II-GFP was not re-internalized to lysosomes, but remained at the surface (See Supplementary Information 1C).

During maturation, dendritic cells exhibit a transient intermediate phenotype in which MHC II molecules are found in peripheral, non-lysosomal class II vesicles^{7,11}. By immunofluorescence of dendritic cells fixed shortly after LPS stimulation, I-E α -GFP could be found together with endogenous I-A molecules in Lamp-negative class II vesicles (Fig. 1d, arrows). In addition, FACS analysis revealed that after LPS stimulation, assembled I-E and I-A molecules appeared at the cell surface with similar kinetics (not shown). Thus, MHC II-GFP was fully functional and appeared to exhibit the same general pattern of intracellular transport as endogenous MHC II in immature and mature dendritic cells.

As a further control, we found that the retroviral expression

system—or expression of the GFP-tagged construct—did not by itself interfere with or induce dendritic cell maturation. Although infected by defective virus, immature dendritic cells upregulated surface expression of endogenous I-A and CD86 normally in response to a maturation signal (Fig. 1e).

Having established that MHC II-GFP molecules form $\alpha\beta$ -dimers that exhibit the expected intracellular distributions, we next asked whether the tagged MHC II molecules could exit from the lysosomes (MHC II compartments) of immature dendritic cells on maturation. Immature dendritic cells expressing MHC II-GFP were visualized by live-cell video confocal microscopy with or without addition of a maturation stimulus. MHC II-GFP-positive structures (presumably lysosomes) in unstimulated dendritic cells remained quiescent for up to 60 min, exhibiting limited saltatory movement in the perinuclear cytoplasm (Fig. 2a; see also Supplementary Information 2A). In contrast, less than 30 min after stimulation by LPS, the lysosomal structures exhibited marked movement and morphological changes, characterized especially by the formation of extensive tubules labelled with MHC II-GFP that extended 5–10 μ m in length (Fig. 2b, left, red arrowheads; see also Supplementary Information 2B). Within 2 h, MHC II-GFP began to appear on the plasma membrane (right). At higher magnification and faster acquisition rates, the tubules were clearly transient structures, separating from their sites of origin and moving centrifugally towards the cell periphery (Fig. 2c; see also Supplementary Information 2C). The behaviour of an individual tubule, highlighted in the boxed area, is shown as individual frames (Fig. 2c, 0–20 s). The three-colour/three-time-point merged image (Fig. 2c) highlights the large degree of tubulation, extension and transport that

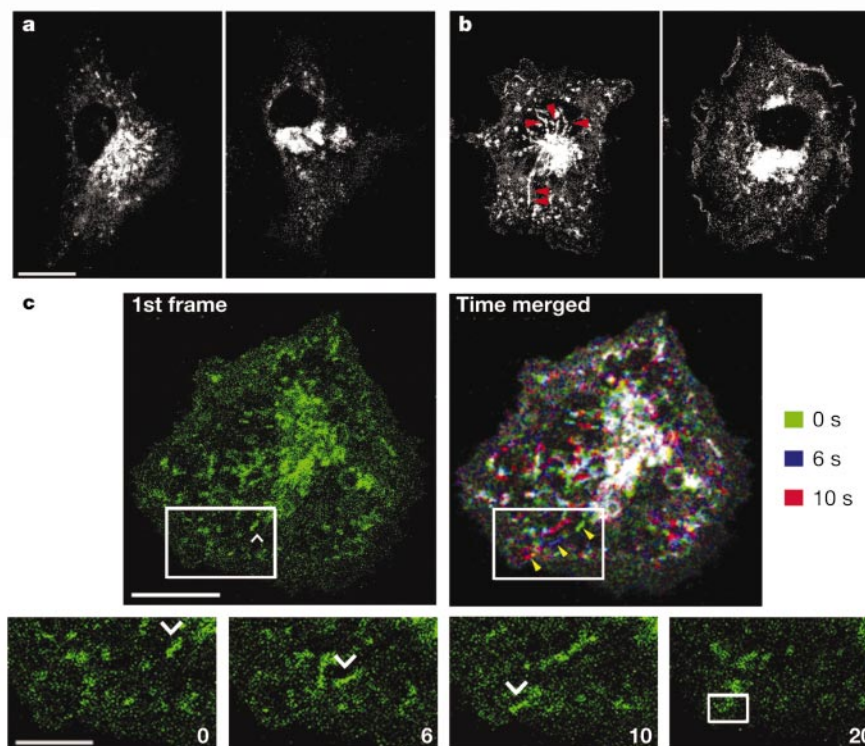


Figure 2 Export of MHC II-GFP molecules from lysosomes to the cell periphery occurs in tubular-vesicular structures in live dendritic cells imaged by confocal microscopy.

a, Unstimulated dendritic cells remain quiescent from the start of live-cell imaging (left) until 4 h later (right). **b**, Stimulated dendritic cells form tubules within 30 min of stimulation (left, red arrowheads; a long tube is highlighted by a double arrowhead) and transport MHC II to the plasma membrane by 4 h (right). Scale bar, 10 μ m. **c**, MHC II-GFP⁺ tubule-like structures in stimulated dendritic cells emanate from a perinuclear region, transit to

the cell periphery, and disappear. The top left panel is an image of a cell corresponding to the first time frame, enlarged in individual frames below. Scale bar, 10 μ m. The top right panel is a time overlay of a cell indicating directionality of MHC II-GFP transport. Structures that do not move over the 10-s period appear white. A tubule moving towards the cell periphery is marked with yellow arrowheads. The bottom panels are labelled with time frame intervals in seconds. A single translocating tubule is labelled with a white arrowhead and disappears at 20 s (white box). Scale bar, 5 μ m.

occurred throughout the cell in just a single 10-s window. Of the 15 unstimulated dendritic cells that were examined, only two formed a limited number of short tubules, in contrast to the more than 50 stimulated cells analysed, all of which exhibited dynamic tubules.

Thus, most of the tubules emanated directly from late endosomes or lysosomes, as demonstrated by dual colour imaging of MHC II-GFP molecules in dendritic cells whose lysosomes had been loaded with Texas red-dextran (Fig. 3a). Formed over a time course of about 30 s, the GFP-labelled tubules were largely devoid of Texas red-dextran suggesting that nascent tubules excluded the bulk of lysosomal contents, presumably owing to their restricted internal volume. These data provide direct confirmation that the class II vesicles observed previously in fixed, maturing dendritic cells were, in fact, lysosome-derived elements⁸.

We performed the same experiment except this time we used cells loaded with Cy3-labelled hen egg lysozyme (HEL), a model antigen that on dendritic-cell maturation contributes peptides that associate with MHC II in lysosomes⁸. As shown in low magnification (Fig. 3b, top; see also Supplementary Information 3B), numerous tubules (arrowheads) extend from the perinuclear area, which was intensely labelled for both MHC II-GFP (green) and Cy3-HEL (red; the large area of overlap is yellow). Many tubules were yellow, suggesting that they contained Cy3-HEL. To better illustrate this point, three consecutive video frames are shown, focusing on a single nascent tubule at higher magnification (Fig. 3b, bottom). The tubule clearly contains both markers. The reasons for the appearance of Cy3-HEL,

but not Texas red-dextran, in the tubules are unclear. It may reflect the ability of HEL to associate with MHC II molecules or the propensity of this cationic protein to bind to the highly anionic lysosomal membrane. In either event, the presence of previously internalized Cy3-HEL in tubules confirmed their lysosomal origin and provided a convenient approach to identifying the origin of any tubule (see below).

Demonstrating that lysosomes can form dynamic and motile tubules after the onset of maturation does not in itself prove that these structures mediate the transfer of MHC II-GFP to the cell surface. To determine directly whether these tubules not only move to the periphery but actually fuse with the plasma membrane, we imaged MHC II-GFP transport in dendritic cells using combined epifluorescence (EPI) and total internal reflectance fluorescence microscopy (TIR-FM)¹⁵. TIR-FM (also known as evanescent wave microscopy) is one of the few techniques that allows a direct observation of fusion, and has been used to image exocytosis of secretory granules, Golgi-derived vesicles and even single synaptic vesicles^{16–19}. In TIR-FM, a beam of light undergoes total internal reflection at the cell–coverslip interface. The light intensity decays exponentially away from the coverslip; hence only a very thin ‘evanescent’ layer (less than 200 nm) is illuminated at the bottom surface of a cell. We first compared TIR-FM images with EPI images collected from the same cell (Fig. 4a). In any single frame, MHC II-GFP molecules imaged by TIR-FM (left) yielded an entirely different pattern than when imaged by EPI (middle). As expected, TIR-FM of dendritic cells expressing MHC II-GFP revealed only those areas of the cell near the coverslip; note the lack of TIR fluorescence in the EPI region where the cell lifts off the coverslip. Bright spots in the TIR-FM image indicate tubular-vesicular structures closely associated or morphologically ‘tethered’ near the cell surface, some of which later fused. As expected, the EPI signal (green) was largely excluded from these regions and instead showed an accumulation of MHC II-GFP (red) in the perinuclear region (merged image, right).

To better illustrate the flux of fluorescence intensity during fusion, a single image of a stimulated dendritic cell was represented using a pseudo-colour scale (Fig. 4b, left; see also Supplementary Information 4B-1, B-2). Candidate fusions were seen all over the attached surface. In total, 25 cells were imaged in these experiments with over 200 fusion events detected (see Methods). One such event is shown over a 36-s time course (Fig. 4b, right). The vesicle/tubule approached the plasma membrane as seen by EPI. As the carriers approached the membrane they were also visible by TIR-FM. During fusion the TIR signal became momentarily brighter (22 s, red arrow) as the vesicle coalesced into the membrane, then dissipated as the membrane protein diffused into the plane of the membrane (23 s, blue arrow). Concomitant with the initial increase of the TIR signal at fusion, the EPI signal diminished, suggesting that fusion had occurred—if the vesicle had not fused and instead detached from the membrane surface, an EPI image would have remained. Although these images suggest a simple diffusive dissipation^{15,19}, closer analysis of the post-fusion diffusion kinetics may reveal maintenance of a preformed complex or cluster of MHC II molecules, as previously suggested on the basis of fixed images⁸. Using TIR-FM with an even shallower penetration depth should, in the future, allow us to investigate this matter more fully as it would substantially improve the signal-to-noise ratio. It is also possible that the clusters observed earlier are not derived from the fusion events but rather form after reaching the surface.

To monitor the origin of the tubules before their final fusion, we performed a combination of TIR-FM with two-colour EPI imaging. This approach allowed us to examine the behaviour of an MHC II-GFP-labelled tubule that could be identified definitively as having come from lysosomes owing to its Cy3-HEL content. Through the use of EPI, video imaging revealed abundant perinuclear lysosomes that were double positive for MHC II-GFP (red) and Cy3-HEL

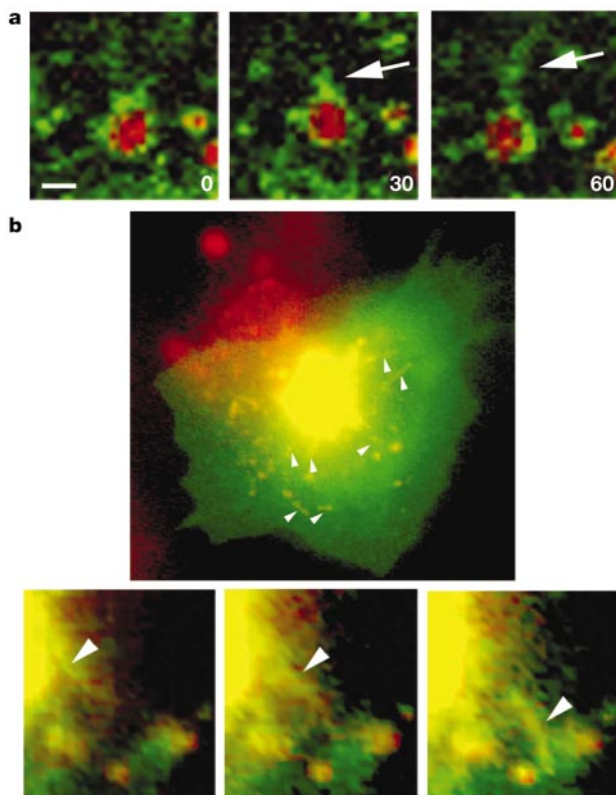


Figure 3 MHC II-GFP tubules originate from lysosomal/late endosomal compartments. **a**, MHC II-GFP⁺ tubules (white arrows) extend from lysosomes loaded with Texas red-dextran over a time course of about 30 s. Panels are labelled with time frame intervals in seconds. Scale bar, 1 μ m. **b**, Stimulated dendritic cells loaded with Cy3-HEL (red) extend several MHC II-GFP (green) tubules, which are also positive for Cy3-HEL and thus appear yellow (white arrowheads). An MHC II-negative cell that took up Cy3-HEL is in the background (upper left) positioned on top of the MHC II-GFP⁺ dendritic cell. Three individual frames (1-s intervals) better illustrate a single double positive tubule (white arrowhead) extending from the perinuclear region.

(blue). In addition, many of the associated tubular elements appeared positive for both markers (Fig. 4c, boxed area and individual frames, yellow arrow). The fusion of a single, double positive Cy3-HEL and MHC II-GFP tubule was illustrated in a series of video frames that separately show the MHC II-GFP TIR-FM signal, the EPI signals for both MHC II-GFP and Cy3-HEL, and a merge of MHC II-GFP TIR-FM (green) and EPI (red) signals (Fig. 4c; 1–16 s; see also Supplementary Information 4C-1, C-2). In the first frame, a single tubule positive for both Cy3-HEL and MHC II-GFP is shown. With time, as the tubule approached the membrane, both EPI signals decreased while the MHC II-GFP TIR-FM signal increased (red arrow marks the maximum level). Fusion occurred by 15 s after which time the signal dissipated (blue arrow). Therefore, by using two-colour EPI combined with TIR-FM imaging we were able to track the formation, transport and ultimate plasma membrane fusion of MHC II-GFP-containing tubules derived from lysosomes.

Thus, dendritic cell lysosomes can initiate a previously unknown retrograde pathway by which selected components can be rescued

from degradation and transferred to the plasma membrane. This tubule-mediated pathway is probably not unique to dendritic cells, but was identified here owing to its synchronous induction on dendritic cell maturation. On the other hand, the capacity for tubule formation may reflect a specialization of dendritic cell lysosomes. There is as yet scant evidence to indicate that MHC II-containing late endocytic compartments in dendritic cells or other antigen-presenting cells comprise a vesicle population physically distinct from conventional late endosomes and lysosomes^{20,21}. The dendritic cell retrograde pathway is clearly distinct from exocytosis of 'secretory lysosomes' in natural killer cells or cytotoxic T lymphocytes²², lysosomes expressing MHC II-GFP in MelJuSo cells²³, or multivesicular bodies (exosomes) in B cells²⁴. These processes are morphologically distinct and also result in the release of bulk lysosomal contents. Although lysosomes are well known to form tubular extensions in macrophages and B cells in response to certain activators^{25,26}, these tubules do not mediate transport to the cell surface but instead may facilitate phagosome-lysosome fusion. The mechanisms that induce tubule formation are unclear. Extension

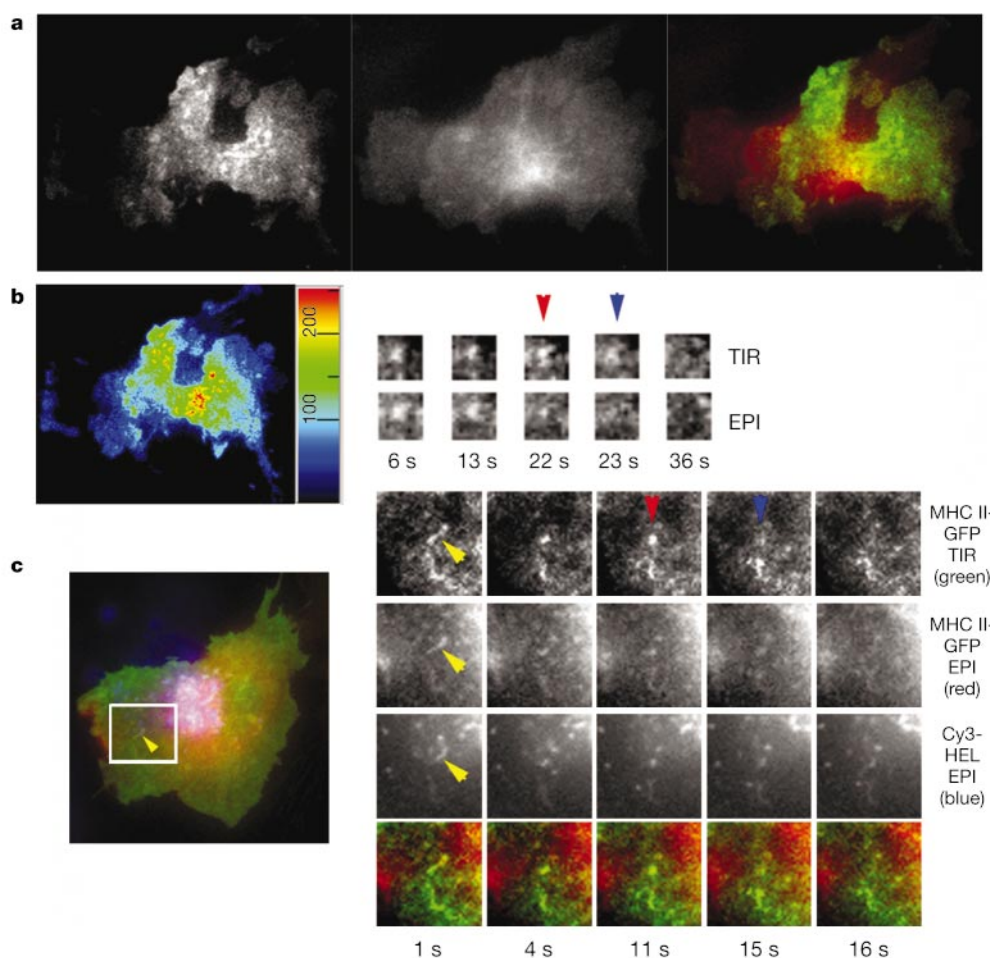


Figure 4 Combined TIR-FM/EPI microscopy demonstrates fusion of lysosome-derived MHC II-GFP structures with the plasma membrane. **a**, TIR-FM (left) illuminates only the cell surface at sites close to the coverslip, whereas EPI (middle) illuminates the entire cell. A merged image (right) highlights differences between the two types of illumination (see text). **b**, Fusion of MHC II-GFP with plasma membrane, indicated by a flash of fluorescence followed by a dissipating signal. The left panel shows a single TIR image represented by a pseudo-colour scale (red, highest fluorescence intensity; blue, lowest fluorescence intensity). Red spots reflect sites of intense membrane activity and candidate fusion events. Supplementary Information 4B illustrates the large number of events that occur. The right panels show sequential frames of a single fusion event simultaneously imaged

by TIR-FM and EPI. Note that the fluorescence signal disappears in the corresponding EPI images (bottom row). **c**, One-colour TIR-FM and two-colour EPI imaging of an MHC II⁺ tubule fusing with the plasma membrane: green, TIR; red, EPI of MHC II-GFP; blue, EPI of Cy3-HEL. The left panel shows a three-colour merge of the first frame highlighting a single tubule that is positive in all three imaging modes. This tubule is enlarged in colour-separated panels showing a 16-s sequence (right). The yellow arrowhead highlights the tubule present in all three channels. The TIR signal was highest at 11 s (red arrowhead) and dissipates at 15 s (blue arrowhead), suggesting a partial fusion event, as the Cy3-HEL signal at 15–16 s does not disappear entirely. The bottom row depicts a merge of MHC II-GFP TIR-FM and MHC II-GFP EPI channels (see Supplementary Information 4C).

along microtubules is probable as is the case for other tubulation events²⁷; however, microtubule disruption does not block the recruitment of MHC II from lysosomes⁸. Also unknown is the degree to which retrograde transport involves selective sorting between MHC II and resident lysosomal membrane components. These and other issues can now be addressed easily using various combinations of recombinant retroviruses together with live cell imaging. □

Methods

Cell culture

Dendritic cells were generated from mouse bone marrow as described previously²⁸. In brief, bone marrow cells from C57BL/6 mice (Jackson Laboratories) were depleted of erythrocytes, T cells, B cells, granulocytes and MHC II-positive cells and subsequently cultured in RPMI (GIBCO) supplemented with 5% FCS, recombinant mouse granulocyte/macrophage-colony stimulating factor (GM-CSF), 20 µg ml⁻¹ gentamycin and 50 µM β-mercaptoethanol (dendritic cell growth media). Recombinant mouse GM-CSF was produced as culture supernatant from J558L cells transfected with the mouse GM-CSF (a gift of D. Gray). Before dendritic cell differentiation, proliferating precursors were infected with retrovirus on day 2 of culture as described previously²⁹. Non-adherent cells were removed carefully and fresh media was added every two days. Immature dendritic cells were imaged on days 4 and 5 of culture. Mature dendritic cells were generated by adding 100 ng ml⁻¹ LPS (Sigma) or bacteria to disaggregated and replated cultures for 24–48 h.

Antibodies

For FACS, staining was performed using the following antibodies: PE-conjugated 14.4.4S, PE-conjugated CD86, PE-conjugated anti-I-A^b and cychrome-conjugated anti-CD11c (all from Pharmingen). Mouse I-Eαβ dimers were detected by 14.4.4S, a mouse anti-mouse monoclonal antibody, and mouse Lamp-2 using the rat anti-mouse monoclonal antibody³. Alexa 594-conjugated goat anti-mouse or goat anti-rat secondary antibodies (Molecular Probes) were used for detection.

Retrovirus generation and transduction

Retrovirus was generated in ΦNX-ecotropic cells as previously described³⁰. I-Eα-GFP was constructed by polymerase chain reaction (PCR)-based mutagenesis (based on Stratagene QuickChange technique), creating a Bgl site at the 3' end of IEα (using the oligonucleotide 5'-CGACAAGGAGCCCTGAGATCTACCTGGAGGTGCGTTAAATGTGC and 3'-GCACATTTAACGACACCTCAGGTAGATCTCAGGCTCCTGTGTCG) and eliminating the stop codon. EGFP (a gift from T. Hughes) was fused in-frame using the Bgl site. The resulting fusion construct I-Eα-EGFP was cloned into LZRS-pBMN using *EcoRI* sites. This viral vector was transfected into ΦNX-ecotropic cells using Eugene (Roche). We obtained stable transfectants after selection in puromycin. Virus was collected in dendritic cell growth medium for 24 h and used to infect dendritic cells. Infection was performed by adapting a previously described method²⁹. Briefly, virus was collected and supplemented with polybrene (8 µg ml⁻¹) and HEPES (10 mM) and added to dendritic cell cultures. Tissue culture plates were spun in a table-top centrifuge at 32 °C, 2,500 r.p.m. (1,200g) for 2 h. Virus was removed and fresh medium was added. Expression was assayed 48 h after infection.

Confocal microscopy

To label lysosomes, day 4–5 immature dendritic cells were loaded with Texas red-dextran (Molecular Probes) at a concentration of 250 µg ml⁻¹ for 30 min, chased with fresh medium for 30 min and plated for 30 min on poly-L-lysine pre-coated number 1.5 coverslips attached to 35-mm dishes (MatTek). Fresh dendritic cell growth medium supplemented with 10 mM HEPES (Gibco) was added to the plate along with LPS or bacteria if the cells were to be stimulated. Cells were imaged on a Zeiss LSM 510 confocal microscope equipped with a × 100 1.4 NA Plan-Apochromat oil-immersion lens. The cell chamber was heated to 37 °C using a Tempcontrol Digital 37-2 device (Warner Instruments). Laser lines at 488 nm and 545 nm were used for excitation of GFP and Texas red, and emissions wavelengths were separated by band pass (505–530 nm) and long pass (585 nm) filters, respectively. To minimize photobleaching, laser power was typically under 2% of maximum and the pinhole size was set to 1.0–1.2. Framescan rates varied from 1.9 to 3.9 s with a line average of 2; multi-tracking (line switching) was used for two-colour imaging. Acquisition was performed using Zeiss LSM 510 version 3.0 software and processing was completed using Adobe Photoshop 5.5 and Openlab 3.0 (Improvision). Movies were assembled using Graphic Converter version 3.8.

TIR-FM and analysis

TIR-FM was achieved using an objective-type set-up. An Olympus IX-70 inverted microscope was fitted with an Olympus × 60 1.45 NA Plan-apochromat TIR-FM oil-immersion lens, mounted on a piezo-electric focus drive (Physics Instruments). Cells were plated on MatTek coverslip dishes as above, and were imaged at 32–35 °C in a heated room. For TIR illumination a 488-nm, 100 mW argon laser (Melles Griot) was electronically shuttered and coupled via single mode fibre (KineFLEX; PointSource) to a dual fibre TIR/monochromator condenser (Till Photonics). TIR illumination used in this configuration had an estimated penetration depth (1/e) of about 93 nm, assuming refractive index $n_{\text{cytosol}} \approx 1.36$ and TIR critical angle $\theta_{\text{max}} \approx 69.7^\circ$ (for calculation see ref. 15). For epi-illumination, we used an electrically shuttered Polychrome IV monochromator (Till

Photonics). Light passed through a dual GFP/dsRed TIR filter block (AHF Analysentechnik) and was detected with an Imago QE-cooled charge-coupled device (CCD) camera (PCO; full chip, 1,376 × 1,040 pixel, 6.45 µm² per pixel, 16 MHz scan rate). Rapid sequential (less than 10-ms delay) TIR and EPI illumination was accomplished using Till Photonics hardware controller and TillVision software. Typically, dual images were acquired at 1–2 Hz with 2 × 2 binning and exposure times of 400 ms and under 50 ms for TIR and EPI illumination, respectively. Post acquisition analysis and processing were done using macros in TillVision software. For analysis of individual fusion events, a 5 × 5 pixel square was centred around the fusion site and the average pixel intensity was measured over time. Fusion events were specifically identified by a 'signature' fusion intensity profile¹⁵ and correlated with loss of signal in both the EPI and TIR channels. By these criterion a fusing vesicle gave a different signal than a fluorescent patch or a vesicle that approaches the cell surface and subsequently retreats into the cell. Note, however, at the current penetration depth of about 100 nm, fusion events are still relatively dim. Fusion events were manually detected and counted while looping 300–6,000 time-lapse frames. Images were exported as TIFF images, processed using Adobe Photoshop version 5.5 and converted into QuickTime movies using Graphic Converter version 3.8.

Received 22 April; accepted 2 July 2002; doi:10.1038/nature01006.

1. Banchereau, J. & Steinman, R. M. Dendritic cells and the control of immunity. *Nature* **392**, 245–252 (1998).
2. Mellman, I. & Steinman, R. M. Dendritic cells: specialized and regulated antigen processing machines. *Cell* **106**, 255–258 (2001).
3. Cella, M., Engering, A., Pinet, V., Pieters, J. & Lanzavecchia, A. Inflammatory stimuli induce accumulation of MHC class II complexes on dendritic cells. *Nature* **388**, 782–787 (1997).
4. Garrett, W. S. et al. Developmental control of endocytosis in dendritic cells by Cdc42. *Cell* **102**, 325–334 (2000).
5. Pierre, P. et al. Developmental regulation of MHC class II transport in mouse dendritic cells. *Nature* **388**, 787–792 (1997).
6. Pierre, P. & Mellman, I. Developmental regulation of invariant chain proteolysis controls MHC class II trafficking in mouse dendritic cells. *Cell* **93**, 1135–1145 (1998).
7. Inaba, K. et al. The formation of immunogenic major histocompatibility complex class II-peptide ligands in lysosomal compartments of dendritic cells is regulated by inflammatory stimuli. *J. Exp. Med.* **191**, 927–936 (2000).
8. Turley, S. J. et al. Transport of peptide-MHC class II complexes in developing dendritic cells. *Science* **288**, 522–527 (2000).
9. Sallusto, F., Cella, M., Danieli, C. & Lanzavecchia, A. Dendritic cells use macropinocytosis and the mannose receptor to concentrate macromolecules in the major histocompatibility complex class II compartment: downregulation by cytokines and bacterial products. *J. Exp. Med.* **182**, 389–400 (1995).
10. West, M. A., Prescott, A. R., Eskelinen, E. L., Ridley, A. J. & Watts, C. Rac is required for constitutive macropinocytosis by dendritic cells but does not control its downregulation. *Curr. Biol.* **10**, 839–848 (2000).
11. Kornfeld, S. & Mellman, I. The biogenesis of lysosomes. *Annu. Rev. Cell Biol.* **5**, 483–525 (1989).
12. Kleijmeer, M. et al. Reorganization of multivesicular bodies regulates MHC class II antigen presentation by dendritic cells. *J. Cell Biol.* **155**, 53–63 (2001).
13. Mathis, D. J., Benoist, C., Williams, V. E., Kanter, M. & McDevitt, H. O. Several mechanisms can account for defective Eα gene expression in different mouse haplotypes. *Proc. Natl Acad. Sci. USA* **80**, 273–277 (1983).
14. Ozato, K., Mayer, N. & Sachs, D. H. Hybridoma cell lines secreting monoclonal antibodies to mouse H-2 and Ia antigens. *J. Immunol.* **124**, 533–540 (1980).
15. Toomre, D., Steyer, J. A., Keller, P., Almers, W. & Simons, K. Fusion of constitutive membrane traffic with the cell surface observed by evanescent wave microscopy. *J. Cell Biol.* **149**, 33–40 (2000).
16. Steyer, J. A. & Almers, W. A real-time view of life within 100 nm of the plasma membrane. *Nature Rev. Mol. Cell Biol.* **2**, 268–275 (2001).
17. Toomre, D. & Manstein, D. J. Lighting up the cell surface with evanescent wave microscopy. *Trends Cell Biol.* **11**, 298–303 (2001).
18. Zenisek, D., Steyer, J. A. & Almers, W. Transport, capture and exocytosis of single synaptic vesicles at active zones. *Nature* **406**, 849–854 (2000).
19. Schmoranz, J., Goulian, M., Axelrod, D. & Simon, S. M. Imaging constitutive exocytosis with total internal reflection fluorescence microscopy. *J. Cell Biol.* **149**, 23–32 (2000).
20. Pierre, P. et al. HLA-DM is localized to conventional and unconventional MHC class II-containing endocytic compartments. *Immunity* **4**, 229–239 (1996).
21. Kleijmeer, M. J., Morkowski, S., Griffith, J. M., Rudensky, A. Y. & Geuze, H. J. Major histocompatibility complex class II compartments in human and mouse B lymphoblasts represent conventional endocytic compartments. *J. Cell Biol.* **139**, 639–649 (1997).
22. Stinchcombe, J. C. & Griffiths, G. M. Regulated secretion from hemopoietic cells. *J. Cell Biol.* **147**, 1–6 (1999).
23. Wubbolts, R. et al. Direct vesicular transport of MHC class II molecules from lysosomal structures to the cell surface. *J. Cell Biol.* **135**, 611–622 (1996).
24. Raposo, G. et al. B lymphocytes secrete antigen-presenting vesicles. *J. Exp. Med.* **183**, 1161–1172 (1996).
25. Lankar, D. et al. Dynamics of major histocompatibility complex class II compartments during B cell receptor-mediated cell activation. *J. Exp. Med.* **195**, 461–472 (2002).
26. Racosin, E. L. & Swanson, J. A. Macropinosome maturation and fusion with tubular lysosomes in macrophages. *J. Cell Biol.* **121**, 1011–1020 (1993).
27. Cole, N. B. & Lippincott-Schwartz, J. Organization of organelles and membrane traffic by microtubules. *Curr. Opin. Cell Biol.* **7**, 55–64 (1995).
28. Inaba, K. et al. Generation of large numbers of dendritic cells from mouse bone marrow cultures supplemented with granulocyte/macrophage colony-stimulating factor. *J. Exp. Med.* **176**, 1693–1702 (1992).
29. De Veerman, M. et al. Retrovirally transduced bone marrow-derived dendritic cells require CD4⁺ T cell help to elicit protective and therapeutic antitumor immunity. *J. Immunol.* **162**, 144–151 (1999).
30. Kinsella, T. M. & Nolan, G. P. Episomal vectors rapidly and stably produce high-titer recombinant retrovirus. *Hum. Gene Therapy* **7**, 1405–1413 (1996).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements

We thank members of the Mellman/Warren laboratory for general support and advice, in particular J. Seemann and L. Pelletier. We also thank J. Kagan, A. Neild and C. Roy for confocal microscopy assistance, L. Zheng and A. Bothwell for help with the retroviral system, and T. Hughes for providing the EGFP complementary DNA. We thank O. Bloom, J. Unternaehrer and J. Chow for critical reading of the manuscript, and Olympus for providing the TIR-FM microscope. We also thank the Ludwig Institute for Cancer Research and the NIH for their support of our work.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.M. (e-mail: ira.mellman@yale.edu).

A chromatin remodelling complex that loads cohesin onto human chromosomes

Mohamed-Ali Hakimi*, Daniel A. Bochar*, John A. Schmiesing†, Yuanshu Dong*, Orr G. Barak*, David W. Speicher*, Kyoko Yokomori† & Ramin Shiekhattar*

* The Wistar Institute, 3601 Spruce Street, Philadelphia, Pennsylvania 19104, USA

† University of California, Irvine, Department of Biological Chemistry, 240D, Med Sci I, Irvine, California 92697-1700, USA

Nucleosomal DNA is arranged in a higher-order structure that presents a barrier to most cellular processes involving protein DNA interactions¹. The cellular machinery involved in sister chromatid cohesion, the cohesin complex, also requires access to the nucleosomal DNA to perform its function in chromosome segregation^{2–10}. The machineries that provide this accessibility are termed chromatin remodelling factors¹¹. Here, we report the isolation of a human ISWI (SNF2h)-containing chromatin remodelling complex that encompasses components of the cohesin and NuRD complexes. We show that the hRAD21 subunit of the cohesin complex directly interacts with the ATPase subunit SNF2h. Mapping of hRAD21, SNF2h and Mi2 binding sites by chromatin immunoprecipitation experiments reveals the specific association of these three proteins with human DNA elements containing Alu sequences. We find a correlation between modification of histone tails and association of the SNF2h/cohesin complex with chromatin. Moreover, we show that the association of the cohesin complex with chromatin can be regulated by the state of DNA methylation. Finally, we present evidence pointing to a role for the ATPase activity of SNF2h in the loading of hRAD21 on chromatin.

We have previously shown that human SNF2h resides in two distinct complexes in HeLa nuclear extract. These are a complex of relative molecular mass (M_r) 670,000 (670K), WCRF/hACF, that eluted in the 1M KCl fraction of the phosphocellulose (P11) chromatography, and a larger complex of M_r 1,500–2,000K eluting in the 0.5-M KCl fraction of P11 (ref. 12). Moreover, immunoprecipitation of SNF2h from the 0.5-M or 1-M KCl eluate of P11 demonstrated the specific association of WCRF180/hACF1 only with the SNF2h in the 1-M P11 fraction (Fig. 1a). To determine the components of the larger SNF2h-containing complex found in the 0.5-M P11 fraction, we purified SNF2h following the scheme in Fig.

1b. Silver stain analysis of the last chromatographic step revealed the association of SNF2h with a complex of approximately 15 polypeptides (Fig. 1c). Western blot analysis of the Superose 6 column fractions indicated that SNF2h eluted with a broad profile encompassing the relative molecular masses of 2,000K to 670K (Fig. 1c). Interestingly, although the two-subunit WCRF/hACF complex could not be detected in the initial steps of the purification, the 670 K SNF2h-containing complex was shown eluting in fractions (33–36) of Superose 6 (Fig. 1c, see the western blot for WCRF180). A combination of ion trap mass spectrometry and western blot analysis identified the other components of the M_r 2,000K complex as subunits of the NuRD chromatin remodelling complex^{13–15}, and the polypeptides involved in the sister chromatid cohesion. All four subunits of the core-cohesin complex—SMC1, SMC3, SA1/SA2 and hRAD21—were identified by mass spectrometry. Western blot analysis confirmed the coelution of hRAD21 and components of the NuRD complex with SNF2h in fractions 27–30 (Fig. 1c). These findings indicate that a fraction of SNF2h displays a chromatographic profile consistent with being a component of a complex that also contains cohesin and the subunits of the NuRD complex.

To rigorously demonstrate that the cohesin-containing SNF2h

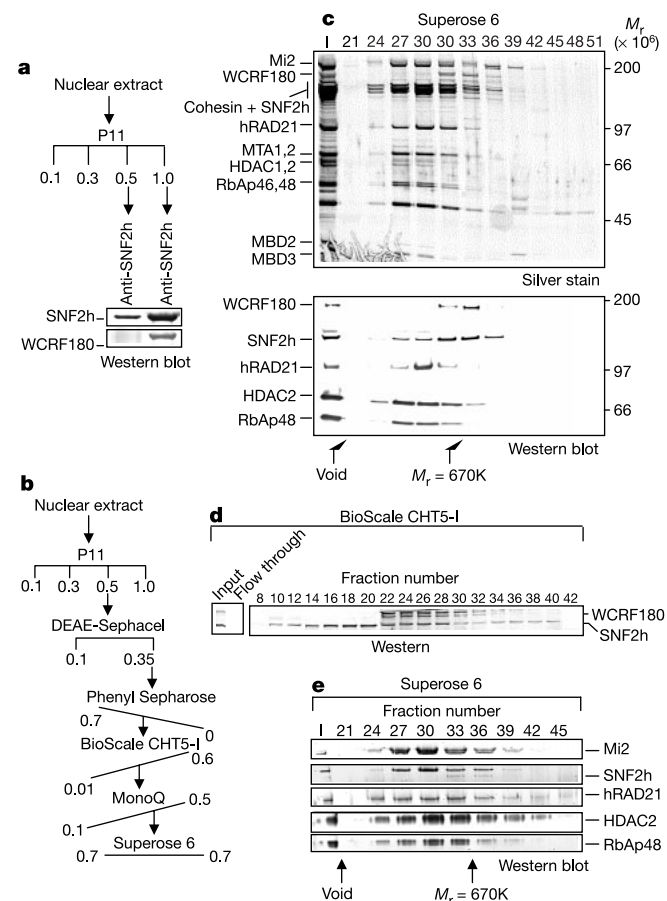


Figure 1 Isolation of an SNF2h complex containing cohesin. **a**, HeLa nuclear extract was fractionated by chromatography and the fractions shown by an arrow were used for affinity-purification followed by western blotting as described²⁴. **b**, Purification scheme. HeLa nuclear extract was fractionated by chromatography as described in Methods. **c**, Silver staining and western blot analysis of Superose 6 fractions (15 μ l). The proteins analysed are indicated to the left of the figure. **d**, Western blot analysis of BioScale CHT5-I column fractions using antibodies shown on the right of the figure. Fractions 18–20 were pooled for analysis on a subsequent Mono Q column. **e**, Western blot analysis of Superose 6 column fractions using antibodies against proteins to the right of the figure. Throughout this purification fractions containing WCRF were excluded.

complex is distinct from the WCRF/hACF complex, we repeated the conventional chromatographic purification following the previous scheme with the exception that WCRF180-containing fractions were excluded from the pool generated from each column (for example in Fig. 1d only fractions 18 through 20 were pooled to be analysed on a subsequent chromatographic step). This procedure yielded a single peak of SNF2h immunoreactivity precisely coeluting with the components of the NuRD and cohesin complex on the final Superose 6 gel filtration (Fig. 1e).

To further establish the stable association of SNF2h with components of the cohesin complex and NuRD, we developed polyclonal antibodies against SNF2h and used them to affinity purify the complex according to the scheme in Fig. 2a. Analysis of the SNF2h affinity eluate using Colloidal blue staining and western blot analysis revealed an identical polypeptide composition to that obtained after conventional purification of SNF2h (Fig. 2b). To establish that the association of the components of the complex is not mediated through DNA, we repeated the affinity purification in

the presence of either ethidium bromide ($300 \mu\text{g ml}^{-1}$) or micrococcal nuclease (MNase, 2 activity units). The association of SNF2h with either the NuRD or cohesin subunits were unaffected by either treatment (Fig. 2b).

The RAD21 protein is an essential subunit of the cohesin complex involved in sister chromatid cohesion^{4–9}. Indeed, proteolytic cleavage of RAD21 at the onset of the metaphase-to-anaphase transition accounts for the separation of sister chromatids at anaphase^{9,16,17}. To further confirm the association of SNF2h with cohesin, we asked whether the anti-hRAD21 antibodies immunoprecipitate the chromatin remodelling complex. As Fig. 2c indicates, anti-hRAD21 antibodies specifically immunoprecipitated SNF2h and Mi2 proteins from the 0.5-M P11 fraction. Similarly, antibodies against HDAC2 specifically immunoprecipitated SNF2h, cohesin, and the components of the NuRD complex (Fig. 2d). These findings demonstrate the stable association of hRAD21 with SNF2h and components of the NuRD complex. Western blot analysis of NuRD, SNF2h and RAD21 throughout the conventional purification (Fig. 2e; see analysis of initial steps) indicates that approximately 15–20% of each component forms a stable SNF2h/NuRD/cohesin complex. To test whether RAD21 and SNF2h can directly associate, each protein was produced as a full-length recombinant and analysed for protein–protein interactions (Fig. 2f). This analysis demonstrated a direct association between hRAD21 and SNF2h (Fig. 2f). Moreover,

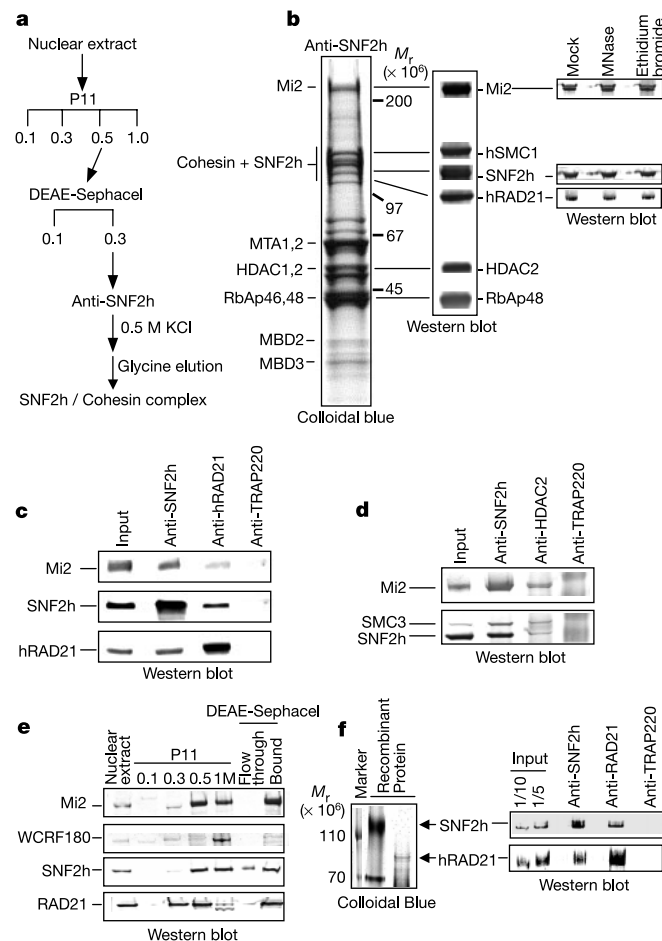


Figure 2 Affinity-purification of the SNF2h/cohesin complex. **a**, Purification scheme for affinity-purification as described in Methods. **b**, Affinity-purification was followed by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) (4–15%), and proteins were visualized by either Colloidal blue staining or western blot analysis in the absence or presence of micrococcal nuclease (MNase) or ethidium bromide. **c**, **d**, Affinity-purification using antibodies denoted on top of the figure followed by western blot analysis using antibodies against proteins to the left of the figure. **e**, Western blot analysis of column fractions (10 μg) using antibodies to the left of the figure. **f**, The recombinant protein Flag-SNF2h and (His)6-hRAD21 were separated in an SDS–PAGE, and proteins were visualized by Colloidal blue staining. Western blotting of the protein–protein assay (described in Methods) was performed using anti-SNF2h and anti-hRAD21 antibodies.

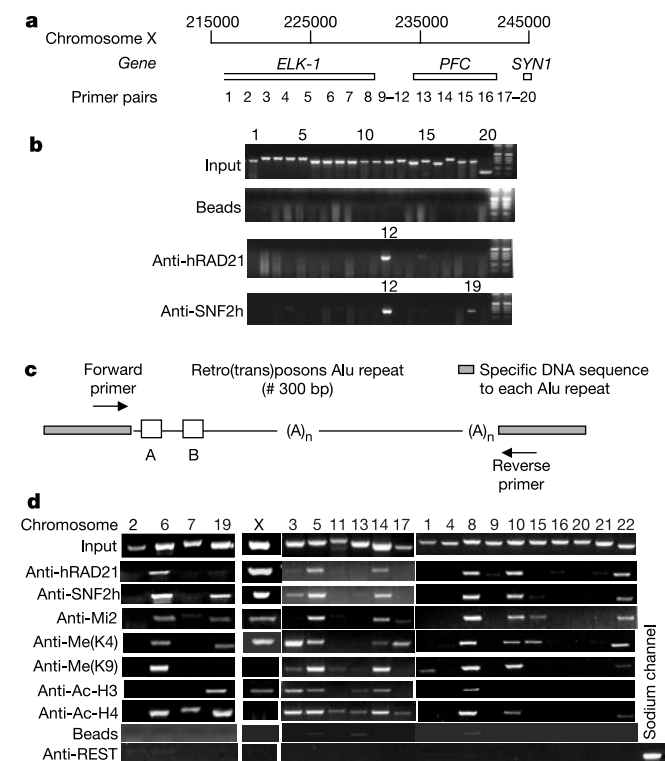


Figure 3 Association of SNF2h/Mi2/cohesin complex with human DNA containing Alu repeats. **a**, A physical map of a chromosome X region showing various sequence features. **b**, 293 cells were processed for chromatin immunoprecipitation (ChIP) as described in Methods using anti-hRAD21 and anti-SNF2h antibodies. A chromatin sample without addition of antibody was used as negative control (beads) and genomic DNA as positive control for the PCR reaction (input). **c**, Diagram of an Alu sequence with the specific primers selected in the flanking regions. A and B represent two conserved Alu boxes; n is the number of poly A stretches. **d**, ChIP analysis of 21 independent Alu repeats on 21 different chromosomes using anti-RAD21, anti-SNF2h, anti-Mi2, anti-methylated histone H3 at lysine 4 or 9 (Upstate Biotechnology), anti-acetylated histone H3, anti-acetylated histone H4 (Upstate Biotechnology) and anti-REST antibodies.

our quantitative analysis indicated that approximately 20% of each recombinant protein forms a stable complex (Fig. 2f). Further protein–protein interaction experiments revealed that SNF2h and SMC1 proteins bind to distinct domains of RAD21 (see Supplementary Information).

Chromosome association sites for the mammalian cohesin complex has not been described. To gain insight into the functional role for the SNF2h and hRAD21 association, we identified the *in vivo* binding sites for both proteins on human chromosomes using chromatin immunoprecipitation (ChIP) experiments. Because there was no prior information regarding binding to human chromatin for either the hRAD21 or SNF2h proteins, we mapped a 30-kilobase (kb) stretch of the X chromosome. Polymerase chain reaction (PCR) primers were used that amplified regions of approximately 350 base pairs (bp) (Fig. 3a). This analysis revealed a single site for association of hRAD21 and chromatin (Fig. 3b, site 12). We note that SNF2h was found to be associated with the same site (site 12). SNF2h was also found to be associated with an additional site near the gene for synapsin (site 19). These findings reveal a specific cohesin-binding site on the X chromosome which is also occupied by SNF2h.

Closer examination of DNA sequence that is bound by hRAD21 (site 12) revealed the presence of the Alu class of short interspersed

DNA elements. These sequences are present in more than 500,000 copies of a 300-bp dimeric sequence and comprise 5–10% of human genomic DNA^{18,19}. We used ChIP to ask whether hRAD21, SNF2h and Mi2 associate with other DNA sites containing Alu repeats by generating specific primers to the flanking regions of twenty-one independent Alu repeats on twenty-one different chromosomes (Fig. 3c). The ChIP analysis demonstrated the specific association of hRAD21, SNF2h and Mi2 with DNA elements containing Alu repeats on chromosomes 6, X, 5, 14, 8, 10 and 22 (Fig. 3d). It is possible that the sequences flanking the Alu repeats may contribute to cohesin binding. However, analysis of the flanking sequences did not yield a specific sequence similarity. Interestingly, chromosome 19, 3 and 15 DNA sequences displayed SNF2h binding but were devoid of hRAD21 binding (Fig. 3d). These findings reveal a subset of DNA elements containing Alu sequences as a binding site for the hRAD21/SNF2h/NuRD complex.

The sequence comparison of DNA elements containing Alu repeats that are bound by hRAD21 to those in which hRAD21 binding was not observed did not yield any obvious difference, so we reasoned that the DNA and/or the nucleosomes in the Alu repeats which are bound by cohesin may be modified differently. To test this we used antibodies against acetylated histones H3 and H4 and methylated histone H3 at lysine 4 or 9 for the ChIP analysis of the 21 Alu repeats. This analysis revealed a correlation between the methylation of lysine 4 in histone H3 and the binding of SNF2h to chromatin (Fig. 3d). We noticed that histones present in the DNA elements containing Alu repeats for which SNF2h binding was observed also displayed either histones H3 or H4 acetylation (Fig. 3d). These findings point to histone modification as an important determinant in association of SNF2h/cohesin/NuRD complex with the DNA elements containing Alu sequences.

Because Alu repeats are unusually rich in CpG dinucleotides which could be a target for the mammalian DNA methyltransferase^{19,20}, we asked whether the state of DNA methylation may govern cohesin binding to DNA elements containing Alu repeats. ChIP experiments were repeated in control cells and cells treated with 5-azacytidine, a DNA methyltransferase inhibitor²¹. Analysis of Alu repeats using a methylase-sensitive restriction enzyme digestion analysis following 5-azacytidine treatment confirmed the decrease in DNA methylation (Fig. 4a). Remarkably, chromosome 3, 15, 17, 19 and 20 DNA sequences, which in control cells only interacted with SNF2h or displayed no binding to either hRAD21 or SNF2h, displayed a strong binding to both SNF2h and hRAD21 following treatment with 5-azacytidine (Fig. 4b). These results implicate a role for DNA methylation in the association of cohesin with chromatin, although this may be an indirect consequence of DNA methylation.

To ask whether the chromatin remodelling activity of SNF2h assisted in association of hRAD21 and chromatin, we examined the effect of expression of SNF2h protein that contains a point mutation (K211R) in the nucleotide-binding motif that abrogates ATP hydrolysis. A similar mutation was shown to disrupt the chromatin remodelling activity of mammalian BRG1 and yeast ISWI proteins^{22,23}. Analysis of the association of hRAD21 with chromatin following transfection of either wild-type SNF2h or mutant SNF2h expression vectors revealed the disruption of hRAD21 binding to DNA elements containing Alu repeats by the mutant SNF2h (Fig. 4c, e; all DNA elements exhibiting cohesin binding were analysed). This effect was specific: not only was hRAD21 binding unaffected by wild-type SNF2h but also the mutant SNF2h did not disrupt the association of SNF2h and chromatin (Fig. 4c, e). Moreover, western blot analysis indicated that both proteins are expressed to the same levels (Fig. 4d). Transfection of mutant SNF2h similarly reduced cohesin binding in DNA elements containing Alu sequences that displayed cohesin association following 5-azacytidine treatment (Fig. 4f). These results reveal a functional association of SNF2h and the cohesin complex.

Thus here we have shown the following: we demonstrate a stable

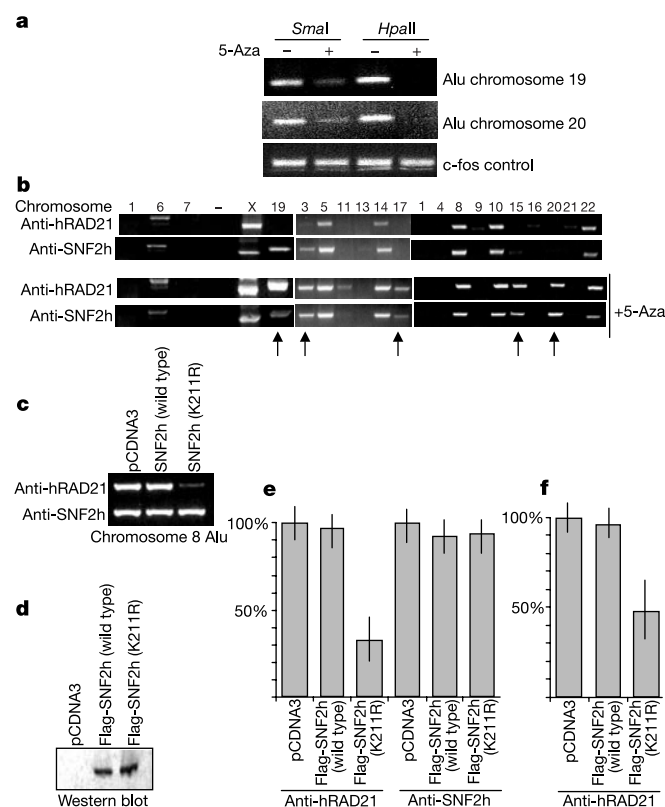


Figure 4 Roles of DNA methylation and chromatin remodelling on cohesin loading. **a**, PCR based methylation assay as described in Methods. **b**, ChIP from either control 293 cells or 293 cells treated with 50 μ M 5-azacytidine for 15 h. Arrows depict the Alu repeats that display increased hRAD21 binding following 5-azacytidine treatment. **c**, After transfection of 293 cells with either pCDNA3, pCDNA3 (Flag-SNF2h wild type) or pCDNA3 (Flag-SNF2h mutant K211R) a representative ChIP on Alu 8 is shown. **d**, Western blot analysis using anti-Flag antibodies following transfection of pCDNA3, pCDNA3 (Flag-SNF2h wild type) or pCDNA3 (Flag-SNF2h mutant K211R). **e**, **f**, ChIP similar to **c** for all Alu repeats display binding for RAD21 before (**e**) or after (**f**) 5-azacytidine treatment in three independent experiments. The data from all sites were pooled in order to graph an average of the binding of hRAD21 and SNF2h after transfection.

association of a chromatin remodelling complex with the machinery of sister chromatid cohesion, cohesin. To our knowledge this is the first demonstration of human cohesin, as well as SNF2h complex, association sites on chromosome arms *in vivo*. We uncover DNA elements containing Alu repeats as cohesin-binding sites on human chromosomes. This work reveals a role for the chromatin remodelling activity of SNF2h in mediating the association of cohesin and chromatin. It also reveals a correlation between the state of histone modification and cohesin association sites. Finally, these results point to a role for DNA methylation in modulating cohesin association with DNA elements containing Alu sequences.

Although our studies point to a role for SNF2h chromatin remodelling activity and the state of DNA methylation in modulating cohesin binding *in vivo*, future *in vitro* studies will determine the precise mechanism of these effects. Taken together, our results extend the scope of the function of the chromatin remodelling machinery beyond that of transcription and DNA repair to encompass association of cohesin with human chromosomes. □

Methods

Conventional chromatographic purification of the SNF2h/cohesin complex

The SNF2h/cohesin complex was purified from 2 g of HeLa nuclear extract (Fig. 1b). Nuclear extract was loaded on a 250-ml column of phosphocellulose (P11, Whatman) and fractionated stepwise by the indicated KCl concentrations in buffer A (20 mM Tris.HCl, pH 7.9, 0.2 mM EDTA, 10 mM β ME, 10% glycerol, 0.2 mM phenylmethyl sulphonyl fluoride (PMSF)). The P11 0.5 M KCl fraction (250 mg) was loaded on a 45-ml DEAE-Sephacel column (Pharmacia) and eluted with 0.35 M KCl. The 0.35 M KCl elution (140 mg) was dialysed to 700 mM NH_4SO_4 in buffer HB (20 mM HEPES, pH 7.6, 4 mM dithiothreitol, 0.5 mM EDTA, 10% glycerol, 0.5 mM PMSF, 1 $\mu\text{g ml}^{-1}$ aprotinin, 1 $\mu\text{g ml}^{-1}$ leupeptin, and 1 $\mu\text{g ml}^{-1}$ pepstatin) and loaded on a Phenyl Superose HR 10/10. The column was resolved using a linear 10 column volume gradient of 700 to 0 mM NH_4SO_4 in buffer HB. SNF2h-containing fractions were dialysed to 10 mM K_2PO_4 in buffer HA (5 mM HEPES, pH 7.6, 1 mM dithiothreitol, 0.5 mM PMSF, 10 μM CaCl₂, 10% glycerol, 40 mM KCl, 1 $\mu\text{g ml}^{-1}$ aprotinin, 1 $\mu\text{g ml}^{-1}$ leupeptin, and 1 $\mu\text{g ml}^{-1}$ pepstatin) and loaded on a BioScale CHT5-I column (BioRad). The column was resolved using a linear 15-column volume gradient of 10 to 600 mM K_2PO_4 in buffer HA. Fractions containing SNF2h were dialysed to 100 mM KCl in buffer A containing 1 $\mu\text{g ml}^{-1}$ aprotinin, leupeptin and pepstatin, and loaded on Mono Q HR 5/5 (Pharmacia). The column was resolved using a linear 10-column volume gradient of 100 to 500 mM KCl in buffer A containing 1 $\mu\text{g ml}^{-1}$ aprotinin, leupeptin and pepstatin. SNF2h-containing fractions were fractionated on a Superose 6 HR 10/30 (Pharmacia) equilibrated in 0.7 M KCl in buffer A containing 0.1% NP-40, and 1 $\mu\text{g ml}^{-1}$ aprotinin, leupeptin and pepstatin. Superose 6 was calibrated using molecular weight standards from Pharmacia. The Void was determined according to manufacturer's guidelines (1/3 of column volume or 7 ml). Estimation of the SNF2h/cohesin complex relative molecular weight was based on the elution profile of a M_r 2,000K SWI/SNF complex²⁴.

Immunoaffinity-purification of SNF2h complex

HeLa nuclear extract (1.2 g) was fractionated according to the protocol described above using P11 and DEAE-Sephacel columns. The 0.3 M DEAE-Sephacel pool was then dialysed to 150 mM KCl in buffer D (20 mM HEPES, pH 7.9, 0.25 mM EDTA, 20% glycerol and 0.1% Tween 20). For affinity purification of SNF2h complex, polyclonal antibodies raised against an amino-terminal peptide (MEEIFDDASPGKQKEIQEPD) were cross-linked using standard techniques²⁴. The affinity purification was performed as described²⁵. A similar protocol was repeated in the presence of either 300 $\mu\text{g ml}^{-1}$ of ethidium bromide or 2 units of micrococcal nuclease²⁶.

In vitro hRAD21 and SNF2h interaction assays

Full-length hRAD21 was produced in bacteria and full-length SNF2h was produced in insect cells as described²⁴. Equal concentration of two proteins (200 ng) were mixed for 4 h in buffer A containing 100 mM KCl and the mixture was purified using anti-SNF2h, anti-hRAD21 or anti-TRAP220 antibodies and washed using buffer A containing 500 mM KCl. Beads were then boiled and analysed by SDS-PAGE followed by western blotting.

Chromatin immunoprecipitation and PCR

Chromatin immunoprecipitations were performed as described²⁵. Primer pairs (melting temperature 55–65 °C) amplifying 100–450-bp fragments were created using published sequences: Chr 2 (GenBank accession numbers given in parentheses: AC011239), Chr 6 (11128486), Chr 7 (AC004593), Chr 19 (AC007787), Chr X (NT_011584), Chr 3 (AC034191), Chr 5 (AC008766), Chr 11 (AC068107.5), Chr 13 (AL359706), Chr 14 (AL139316.5), Chr 17 (AC005919), Chr 1 (AL513329), Chr 4 (AC096759), Chr 8 (AC090814), Chr 9 (AC007172), Chr 10 (AC079844), Chr 15 (AC090514), Chr 16 (AC010542), Chr 20 (AL031672), Chr 21 (AP001627) and Chr 22 (Z95115). For each primer pair the optimal magnesium concentration (1–2.5 mM MgCl_2) was determined. For individual primer pairs the annealing temperature and number of cycles were adjusted until no signal was detected for the mock-immunoprecipitated DNA ('beads'), but the

amplification on the genomic template was not altered ('input'). Signals obtained with the antibody-immunoprecipitated DNA under these conditions were considered significant. The amplified DNA was separated on 2% agarose. Digital images of ethidium-bromide-stained gels were quantified using KODAK 1D 3.5.3 software.

PCR methylation assay

The experiments were performed as described previously²⁷. 293 cells were treated with or without 50 mM of 5-azacytidine for 24 h, and genomic DNA was isolated and digested with *Sma*I or *Hpa*II. 2 ng of digested DNA were then amplified for 32 cycles by PCR using primers for specific Alu sequences.

Received 16 April; accepted 17 July 2002; doi:10.1038/nature01024.

- Kornberg, R. D. & Lorch, Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**, 285–294 (1999).
- Larionov, V. L., Karpova, T. S., Kouprina, N. Y. & Jouravleva, G. A. A mutant of *Saccharomyces cerevisiae* with impaired maintenance of centromeric plasmids. *Curr. Genet.* **10**, 15–20 (1985).
- Strunnikov, A. V., Larionov, V. L. & Koshland, D. SMC1: an essential yeast gene encoding a putative head-rod-tail protein is required for nuclear division and defines a new ubiquitous protein family. *J. Cell Biol.* **123**, 1635–1648 (1993).
- Michaelis, C., Ciosk, R. & Nasmyth, K. Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell* **91**, 35–45 (1997).
- Guacci, V., Koshland, D. & Strunnikov, A. A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in *S. cerevisiae*. *Cell* **91**, 47–57 (1997).
- Birkenbihl, R. P. & Subramani, S. Cloning and characterization of rad21, an essential gene of *Schizosaccharomyces pombe* involved in DNA double-strand-break repair. *Nucleic Acids Res.* **20**, 6605–6611 (1992).
- Birkenbihl, R. P. & Subramani, S. The rad21 gene product of *Schizosaccharomyces pombe* is a nuclear, cell cycle-regulated phosphoprotein. *J. Biol. Chem.* **270**, 7703–7711 (1995).
- Tatebayashi, K., Kato, J. & Ikeda, H. Isolation of a *Schizosaccharomyces pombe* rad21ts mutant that is aberrant in chromosome segregation, microtubule function, DNA repair and sensitive to hydroxyurea: possible involvement of Rad21 in ubiquitin-mediated proteolysis. *Genetics* **148**, 49–57 (1998).
- Tomonaga, T. *et al.* Characterization of fission yeast cohesin: essential anaphase proteolysis of Rad21 phosphorylated in the S phase. *Genes Dev.* **14**, 2757–2770 (2000).
- Schmiesing, J. A. *et al.* Identification of two distinct human SMC protein complexes involved in mitotic chromosome dynamics. *Proc. Natl Acad. Sci. USA* **95**, 12906–12911 (1998).
- Fyodorov, D. V. & Kadonaga, J. T. The many faces of chromatin remodeling: SWITCHing beyond transcription. *Cell* **106**, 523–525 (2001).
- Bochar, D. A. *et al.* A family of chromatin remodeling factors related to Williams syndrome transcription factor. *Proc. Natl Acad. Sci. USA* **97**, 1038–1043 (2000).
- Tong, J. K., Hassig, C. A., Schnitzler, G. R., Kingston, R. E. & Schreiber, S. L. Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature* **395**, 917–921 (1998).
- Xue, Y. *et al.* NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol. Cell.* **2**, 851–861 (1998).
- Zhang, Y., LeRoy, G., Seelig, H. P., Lane, W. S. & Reinberg, D. The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* **95**, 279–289 (1998).
- Waizenegger, I. C., Hauf, S., Meinke, A. & Peters, J. M. Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. *Cell* **103**, 399–410 (2000).
- Uhlmann, F., Lottspeich, F. & Nasmyth, K. Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc 1. *Nature* **400**, 37–42 (1999).
- Mighell, A. J., Markham, A. F. & Robinson, P. A. Alu sequences. *FEBS Lett.* **417**, 1–5 (1997).
- Schmid, C. W. Does SINE evolution preclude Alu function? *Nucleic Acids Res.* **26**, 4541–4550 (1998).
- Hellmann-Blumberg, U., Hintz, M. F., Gatewood, J. M. & Schmid, C. W. Developmental differences in methylation of human Alu repeats. *Mol. Cell Biol.* **13**, 4523–4530 (1993).
- Broday, L., Lee, Y. W. & Costa, M. 5-azacytidine induces transgene silencing by DNA methylation in Chinese hamster cells. *Mol. Cell Biol.* **19**, 3198–3204 (1999).
- Khavari, P. A., Peterson, C. L., Tamkun, J. W., Mendel, D. B. & Crabtree, G. R. BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. *Nature* **366**, 170–174 (1993).
- Gelbart, M. E., Rechsteiner, T., Richmond, T. J. & Tsukiyama, T. Interactions of Isw2 chromatin remodeling complex with nucleosomal arrays: analyses using recombinant yeast histones and immobilized templates. *Mol. Cell Biol.* **21**, 2098–2106 (2001).
- Bochar, D. A. *et al.* BRCA1 is associated with a human SWI/SNF-related complex: linking chromatin remodeling to breast cancer. *Cell* **102**, 257–265 (2000).
- Hakimi, M. A. *et al.* A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc. Natl Acad. Sci. USA* **99**, 7420–7425 (2002).
- Lai, J.-S. & Herr, W. Ethidium bromide provides a simple tool for identifying genuine DNA-independent protein associations. *Proc. Natl Acad. Sci. USA* **89**, 6958–6962 (1992).
- Stoger, R. *et al.* Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. *Cell* **73**, 61–71 (1993).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements

We thank W. Wang and G. Mandel for the gift of Mi2 and REST antibodies, respectively. We also thank T. Nagase at the Kazusa DNA Research Institute in Japan for providing the hRAD21 cDNA. We also thank H. C. Gregson and A. R. Ball, Jr for analysis of Rad21 and SNF2h interactions. This work was supported by a grant from NIH to R.S. and in part by a March of Dimes Basil O'Connor Scholarship and the NIH to K.Y. K.Y. is a Scholar of the

Leukemia & Lymphoma Society. M.-A.H. was supported by a postdoctoral fellowship from Association pour la Recherche sur le Cancer (FRANCE). D.A.B. is a recipient of an NIH training grant.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.S. (e-mail: shiekhattar@wistar.upenn.edu).

Reverse engineering of the giant muscle protein titin

Hongbin Li*, Wolfgang A. Linke†, Andres F. Oberhauser*, Mariano Carrion-Vazquez*, Jason G. Kerkvliet*, Hui Lu‡, Piotr E. Marszalek* & Julio M. Fernandez*

* Department of Physiology and Biophysics, Mayo Foundation, Rochester, Minnesota 55905, USA

† Institute of Physiology and Pathophysiology, University of Heidelberg, D-69120 Heidelberg, Germany

‡ Donald Danforth Plant Science Center, St Louis, Missouri 63132, USA

Through the study of single molecules it has become possible to explain the function of many of the complex molecular assemblies found in cells^{1–5}. The protein titin provides muscle with its passive elasticity. Each titin molecule extends over half a sarcomere, and its extensibility has been studied both *in situ*^{6–10} and at the level of single molecules^{11–14}. These studies suggested that titin is not a simple entropic spring but has a complex structure-dependent elasticity. Here we use protein engineering and single-molecule atomic force microscopy¹⁵ to examine the mechanical components that form the elastic region of human cardiac titin^{16,17}. We show that when these mechanical elements are combined, they explain the macroscopic behaviour of titin in intact muscle⁶. Our studies show the functional reconstitution of a protein from the sum of its parts.

Individual titin molecules span both the A-band and I-band regions of muscle sarcomeres. The I-band part of titin has been identified as the region that is functionally elastic. We study the shortest titin isoform, the N2B isoform found in cardiac-muscle sarcomeres. The elastic I-band region of N2B-titin can be subdivided into four structurally distinct regions (Fig. 1): a proximal immunoglobulin region containing 15 tandem immunoglobulin-like (Ig) domains; a middle N2B segment that contains a 572-residue amino-acid sequence of unknown structure; a 186-amino-acid-long segment rich in proline (P), glutamate (E), valine (V) and lysine (K) residues, named the PEVK region; and a distal Ig region that contains 22 tandem Ig modules¹⁷. We use polyprotein engineering^{18,19} and single-molecule force spectroscopy to dissect the individual mechanical elements of the I-band of cardiac titin and reconstruct the elasticity of cardiac muscle. Polyproteins, when mechanically stretched by single-molecule atomic force microscopy (AFM) give distinctive mechanical fingerprints as their modules unfold sequentially (sawtooth patterns in the force–extension curve)¹⁸, and can be used to positively identify the mechanical features of a single molecule^{19–21} (Supplementary Information).

The top trace in Fig. 1a shows a typical sawtooth pattern measured by stretching a protein composed of eight modules from the proximal tandem Ig region, I4 to I11. The sawtooth pattern shows that all modules unfold in the range of 150–200 pN. However, there is a slight tendency for the first unfolding event to occur at a lower force than later unfolding events. In order

to examine this tendency, we plot the average value of all first unfolding peaks, second peaks, and so on (Fig. 1b, filled circles). A linear fit to the data (Fig. 1b, thin line through filled circles) showed only a weak hierarchy of 12 pN per force peak. Polyproteins constructed using modules I4 (I4₈) and I5 (I5₈) showed similar unfolding forces of 150–200 pN (Fig. 1b, open circles). Hence, it seems that the proximal tandem Ig region has modules of similar mechanical stability. We studied the I4 polyprotein in more detail following the AFM protocols of ref. 18, and measured an unfolding rate of $3 \times 10^{-3} \text{ s}^{-1}$ and a folding rate of 0.33 s^{-1} .

Similar experiments done with polyproteins from the distal Ig region revealed a very different picture. Stretching a protein composed of eight modules from the distal tandem Ig region, I27 to I34, showed a much broader range of unfolding forces, from ~150 pN up to 330 pN (Fig. 1a, bottom trace). As before, we plot the average value of all first unfolding peaks, second peaks, and so on (Fig. 1b, filled squares). A linear fit to the data (Fig. 1b, thin line through

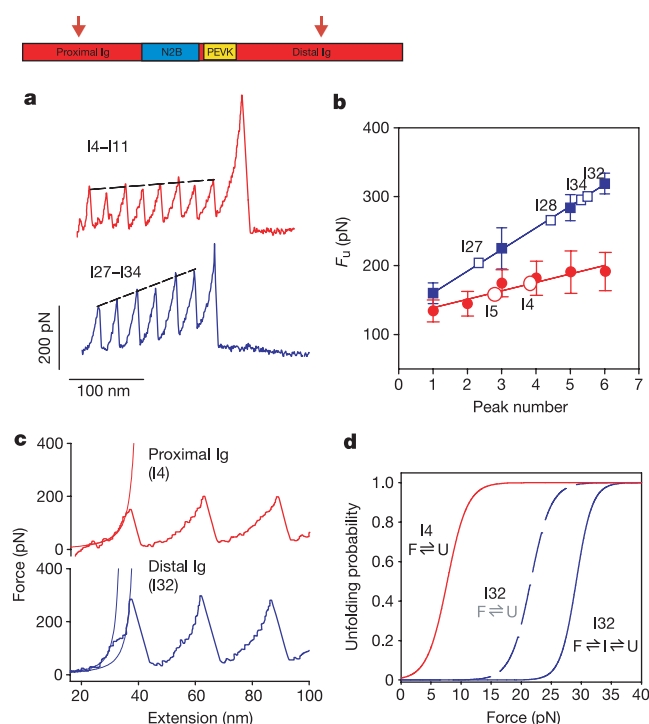


Figure 1 The proximal and distal tandem Ig regions of cardiac titin have different mechanical properties. Inset, the structurally distinct elements of I-band titin. The arrows point to the tandem Ig regions. **a**, Top trace: force–extension curve obtained from an engineered protein comprising domains I4 to I11 of the proximal tandem Ig region. Bottom trace: force–extension curve obtained from a protein comprising domains I27 to I34 of the distal tandem Ig region. **b**, Unfolding forces (F_u) measured for consecutive unfolding peaks (1–6) in AFM recordings of the I4–I11 protein (filled circles) and the I27–I34 protein (filled squares). Recordings obtained from polyproteins containing only I27, I28, I32, or I34 Ig domains (open squares; I27₈: $204 \pm 26 \text{ pN}$, $n = 266$; I28₈: $257 \pm 27 \text{ pN}$, $n = 245$; I34₈: $281 \pm 44 \text{ pN}$, $n = 32$; I32₈: $298 \pm 24 \text{ pN}$, $n = 132$) show a strong hierarchy. The stability of I4 and I5 polyproteins (open circles, I4₈ and I5₈; I4: $171 \pm 26 \text{ pN}$, $n = 136$; I5: $155 \pm 33 \text{ pN}$, $n = 196$) confirms the weak hierarchy of the proximal region. **c**, Top trace: force–extension relationship of an I4 polyprotein (I4₈). The initial part of the force trace, before the first unfolding peak, is well described by the WLC model (thin line). Bottom trace: force–extension relationship for an I32 polyprotein (I32₈) from the distal tandem Ig region of titin. In the initial rising phase of the force–extension curve, a prominent 'hump' appears, indicating the presence of an unfolding intermediate²⁴. **d**, Plot of the steady-state unfolding probability of the I4 and I32 modules as a function of force. I4 is calculated as a simple two-state unfolding system (solid red line). The I32 module is calculated both in the presence (solid blue line) and in the absence (dashed blue line) of the unfolding intermediate.

filled squares) gave a slope of 31.5 pN per force peak. These results indicate a mechanical hierarchy among these modules. In order to determine the mechanical stability of the individual modules and their ordering in the hierarchical unfolding, we constructed several polyproteins: I27₈, I28₈, I32₈ and I34₈. The average unfolding forces were found to be 204 pN for I27 (ref. 18), 257 pN for I28 (ref. 19), 298 ± 24 pN ($n = 132$) for I32 and 281 ± 44 pN ($n = 32$) for I34 (Fig. 1b, open squares). These results contrast with those for the proximal region where no obvious mechanical hierarchy was observed.

Several models of polymer elasticity have been developed to predict the mechanical behaviour of a polymer. As before, we use the worm-like chain (WLC)²² model to fit the force–extension curves of a polyprotein¹⁸. A close examination of the force–extension curve obtained from a proximal Ig module (I4₈, Fig. 1c) shows that the WLC model fits well the force–extension curve preceding each unfolding event (thin red line in Fig. 1c). The unfolding event that occurs is an all-or-none process that can be easily described by a two-state model of the type $F \rightleftharpoons U$ with rate constants for unfolding, $\alpha_u(F)$, and folding, $\beta_f(F)$, that are force dependent²³. Under a constant force F , the probability of unfolding is given by $P_u(F) = \alpha/(\alpha + \beta)$, which has a sigmoidal shape when plotted against the stretching force (Fig. 1d). The plot shows that for I4, $P_u(F) = 0.5$ at a force of 7.7 pN. This result is similar for I5, and is likely to be similar for the other modules of the proximal Ig region.

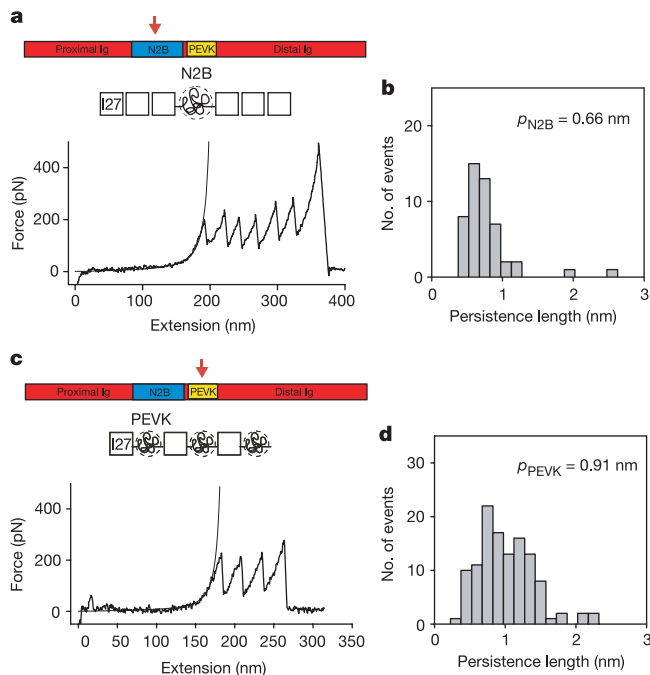


Figure 2 Single-molecule AFM measurements of the mechanical properties of the N2B and PEVK regions of titin. **a**, Top inset: the arrow points to the location of the N213 region in the I-band. Force–extension curve of a protein chimera containing the cardiac N2B unique sequence flanked on either side by three I27 domains (I27₃-N2B-I27₃), bottom inset. A Levenberg–Marquardt fit of the WLC equation (thin line) to the force–extension curve before the first I27 unfolding event measured the contour length, L_c , and persistence length, p , of N2B. **b**, Frequency histogram of persistence-length values. A narrow distribution is found, centred at 0.66 nm. **c**, Top inset: the arrow points to the location of the PEVK region in the I-band. Force–extension curve of a protein chimera containing human cardiac PEVK domains alternating with Ig I27 domains, (I27-PEVK)₃, bottom inset. As in **a**, we used Levenberg–Marquardt fits of the WLC equation to measure L_c and p of the PEVK region (thin line). **d**, Frequency histogram of persistence-length values measured for the PEVK domain. A relatively broad distribution is seen ($p = 0.4$ – 2.5 nm; average value, 0.91 nm).

The WLC model does not fit the force–extension curve of the I32 polyprotein (Fig. 1c, bottom trace) because of a pronounced ‘hump’ that corresponds to an unfolding intermediate before full unfolding²⁴. We have observed a similar intermediate in all of the distal Ig modules tested, whereas we have not observed such an intermediate in the proximal domains. This unfolding intermediate may serve as a kinetic trap to stabilize the distal domains and protect them against unfolding. To illustrate this point, we first ignore the unfolding intermediate, and consider a simple two-state unfolding reaction with $P_u(F) = \alpha/(\alpha + \beta)$. The rate constants, α and β , are calculated from the peak unfolding forces and their dependence on the rate of stretching¹⁸, ignoring the unfolding intermediate. Figure 1d (dashed line) shows that $P_u(F) = 0.5$ at 21.6 pN for I32 in the absence of an unfolding intermediate. When the intermediate is considered, we use a simplified three-state model like $F \rightleftharpoons I \rightleftharpoons U$. Two sets of rate constants describe this model: α_u and β_u corresponding to the main unfolding reaction taken to occur between the intermediate and the unfolded state, and α_i and β_i describing the forward and backward rates of transition to the intermediate state. These last two rate constants were estimated from the data obtained for the intermediate unfolding state of the I27 module²⁴. The unfolding probability for the three-state model is given by:

$$P_u^I(F) = \frac{\alpha_i \alpha_u}{\alpha_i \alpha_u + \beta_i \beta_u + \alpha_i \beta_u} \quad (1)$$

The three-state unfolding probability is a sigmoidal function that is shifted to the right of that calculated without the unfolding intermediate. In this case, $P_u^I = 0.5$ at 29.2 pN. As the on-rate of the unfolding intermediate $\beta_i = 100 \text{ s}^{-1}$ is much faster than the off-rate of the main unfolding event $\alpha_u = 0.01 \text{ s}^{-1}$, the module under force will not go directly to the unfolded state but rather go back to the folded state. Thus, this unfolding intermediate acts as an absorbing state (or buffering state) that kinetically prevents the module from unfolding. This difference in mechanical stability between distal and proximal Ig domains reflects the mechanical topology of these two classes of Ig modules (Supplementary Information).

In order to study the mechanical properties of the N2B segment, we constructed a polyprotein composed of a single N2B module

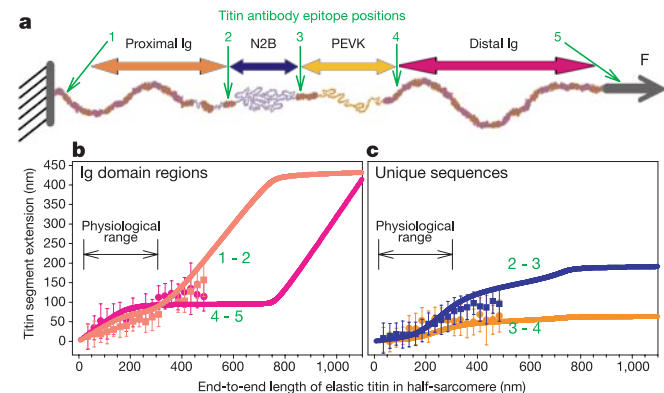


Figure 3 Single-molecule data explain the extensibility of the individual titin segments measured *in situ*. **a**, Schematic diagram showing the four main segments that contribute to the elasticity of titin in the half-sarcomere of cardiac muscle (horizontal arrows). Numbers 1–5 indicate the epitope positions of titin antibodies used to measure the extension of these segments *in situ*. The epitopes move relative to one another when the muscle fibres are stretched. **b**, **c**, Extension of the individual titin segments plotted as a function of the end-to-end extension of I-band titin (symbols). The solid lines were calculated using equations (3)–(5), and the single-molecule data from Table 1. The numbers label the extension of the corresponding epitopes marking the proximal (1–2) and distal (4–5) Ig-domain regions, the N2B unique sequence (2–3) and the PEVK domain (3–4).

Table 1 Mechanical parameters describing the I-band region of human cardiac titin.

	Unfolding rate α (s ⁻¹)	Folding rate β (s ⁻¹)	Persistence length p (nm)	Kuhn length l_k (nm)	Unfolding distance Δx_u (nm)	Folding distance Δx_f (nm)
Proximal Ig domains	$\alpha_u = 3.3 \times 10^{-3}$	$\beta_u = 0.33$	10 (folded) 0.66 (unfolded)	20 (folded) 1.32 (unfolded)	0.25	2.2
N2B unique sequence	—	—	0.66	1.32	—	—
PEVK segment	—	—	0.91	1.82	—	—
Distal Ig domains	$\alpha_u = 8 \times 10^{-5}$	$\beta_u = 1.2$	10 (folded) 0.66 (unfolded)	20 (folded) 1.32 (unfolded)	0.25	2.2
Unfolding intermediate (I), distal Ig domains	$\alpha_i = 1.0 \times 10^{-2}$	$\beta_i = 10^2$	—	—	0.33	0.33

All values were obtained from single molecule force spectroscopy measurements, except for the persistence length of the folded Ig-domain regions, which was measured from EM images (Supplementary Information).

flanked on either side by three tandem I27 domains (I27₃-N2B-I27₃; Fig. 2a), where the I27 modules are used to create a mechanical fingerprint. We used single-molecule AFM to obtain force–extension curves from this polyprotein. We collected 48 recordings like the one in Fig. 2a showing a long initial region, without any unfolding peaks, followed by a sawtooth pattern with four to six consecutive unfolding events. The observed unfolding peaks of ~200 pN spaced by ~28 nm correspond to the characteristic fingerprint of the I27 module^{18–21}. If we observe at least four I27 unfolding events, then N2B must have been stretched when pulling this protein. Given that the extension of the segments of the protein will be hierarchical, from least stable to most stable^{19,20}, the long but featureless part of the trace preceding the sawtooth pattern must correspond to the extension of N2B. Then, extension of N2B occurs at low force and without significant energy barriers limiting its extensibility. This result suggests that the N2B segment has the mechanical properties of a random coil.

The WLC model (thin line, Fig. 2a) fits the force–extension curve of N2B and measures a contour length of 209 nm and a persistence length of $p = 0.74$ nm. Similar measurements made in 48 different recordings gave a distribution of persistence lengths that averaged $p_{N2B} = 0.66$ nm (Fig. 2b). We also measured an average contour length of 232 nm, which agrees well with the expected length of a 572-amino-acid-long polypeptide. We also constructed a polyprotein made of three repeats of the dimer PEVK-I27 (Fig. 2c, ref. 20). WLC fits to the force–extension curve of PEVK (thin line, Fig. 2c) measured an average contour length of 68 nm per PEVK segment²⁰. The persistence length of PEVK varied from 0.4 nm up to 2.5 nm with an average value of $p_{PEVK} = 0.91$ nm (Fig. 2d, ref. 20), suggesting that PEVK could show mechanical conformations that, while still corresponding to a random coil, had different flexibility.

The extensibility of each elastic segment of cardiac I-band titin has been measured in intact cardiac muscle fibres^{6,25}, by following the relative position of several sequence-specific titin antibodies (Fig. 3a). We reconstituted the extensibility of I-band titin by calculating the extension of each segment (proximal Ig, N2B, PEVK, distal Ig) at a given force, and repeating this calculation for a range of forces from 0 up to 40 pN. As all segments experience the same force at all times, the segments extend independently of each other and thus their contributions to the overall length are additive. The total end-to-end length of I-band titin, $x(F)_{I\text{-band}}$, is then calculated as the sum of the extension of all segments. The extensibility of a segment has two components: the entropic spring behaviour and module unfolding, if any^{11–13}.

The N2B and PEVK segments are entropic springs that do not show any unfolding events. These segments are simply modelled by the WLC approach, although with different persistence lengths (we use the average persistence length in each case; $p_{N2B} = 0.66$ nm and $p_{PEVK} = 0.91$ nm). However, the use of the WLC model in this reconstruction is inconvenient, because it gives the force that results from a given extension, $F(x)$, whereas we want to calculate the extension that results from an applied force. The freely jointed chain

model of polymer elasticity²⁶ is described by equation (2).

$$x_{FJC}(F) = L_c u\left(\frac{Fl_k}{k_B T}\right) \quad (2)$$

where L_c is the contour length, $u(Fl_k/k_B T)$ is the Langevin function (where K_B is the Boltzmann constant) and where the Kuhn length $l_k = 2p$ (ref. 22). We can now calculate the extension of these segments for a given force: $x(F)_{N2B}$ and $x(F)_{PEVK}$.

The extensibility of the proximal and distal tandem Ig domain segments, $x(F)_{\text{proximal}}$ and $x(F)_{\text{distal}}$, is also described by equation (2). However, in this case, L_c and l_k depend on module unfolding (Supplementary Information). Thus, the extension of the proximal Ig region under an applied force, $x(F)_{\text{proximal}}$, is fully described by the following three equations:

$$x(F)_{\text{proximal}} = L_c^{\text{folded}}(F) u\left(\frac{F l_k^{\text{folded}}}{k_B T}\right) + L_c^{\text{unfolded}}(F) u\left(\frac{F l_k^{\text{unfolded}}}{k_B T}\right) \quad (3)$$

$$L_c^{\text{folded}}(F) = N(1 - P_u(F))4.4 \quad (4)$$

$$L_c^{\text{unfolded}}(F) = N P_u(F)32.5 \quad (5)$$

where N is the total number of Ig modules in the segment and $P_u(F)$ is the probability of unfolding at a given force; equations (4) and (5) give lengths in units of nm. The extension of $x(F)_{\text{distal}}$ is calculated similarly but including a term for the contribution of the unfolding intermediate. We now calculate the total extension of I-band titin as:

$$x(F)_{I\text{-band}} = x(F)_{\text{proximal}} + x(F)_{N2B} + x(F)_{PEVK} + x(F)_{\text{distal}} \quad (6)$$

for forces ranging from 0 to 40 pN. This calculation creates a table of values relating $x(F)_{I\text{-band}}$ with $x(F)_{\text{proximal}}$, $x(F)_{N2B}$, $x(F)_{PEVK}$ and $x(F)_{\text{distal}}$. We can now compare the extensibility of each I-band titin segment with the extensibility measured *in situ*. The parameters used such as the values of persistence length and the unfolding/folding rate constants correspond to the experimentally determined values listed in Table 1. There are no free parameters in this computation.

Figure 3b compares the extensibility of the tandem Ig regions (proximal, orange line; distal, violet line), calculated with equation (6), with their *in situ* extensibility (symbols). The calculated extensibility of these segments agrees well with the myofibril data. The proximal domains extend first in a fully folded configuration. Unfolding of the proximal region becomes obvious at $x(F)_{I\text{-band}} > 300$ nm, whereas unfolding of the distal Ig region does not occur until much later at $x(F)_{I\text{-band}} > 800$ nm. Hence, the single-molecule data predict that in the physiological range ($0 < x(F)_{I\text{-band}} < 300$ nm) the distal Ig region will never unfold any of its modules whereas the proximal region may see a few of its modules unfold towards the high end of the physiological range. Figure 3c shows plots of $x(F)_{N2B}$ (blue line) and of $x(F)_{PEVK}$ (orange line) versus the end-to-end length of the I-band titin, $x(F)_{I\text{-band}}$. The figure shows

that the calculated extension of these segments fits those measured *in situ* (symbols).

Figure 4 plots the relationship between force and the end-to-end length of I-band titin, $x(F)_{\text{I-band}}$ calculated from equation (6) (solid red line). We compare this calculation with the passive force versus sarcomere length relationship of an intact cardiac myofibril⁶ measured under quasi-steady-state conditions (no viscous or viscoelastic forces present)^{6,27}. The filled symbols in Fig. 4 correspond to single cardiac myofibril data scaled by the number of titin molecules per cross-sectional area of muscle (assumed to be 6×10^9 titin molecules per mm^2)²⁸. The figure shows that the force–extension relationship calculated from the single-molecule AFM data faithfully predicts the force–extension relationship measured in intact myofibrils. So by scaling the single-molecule data, it is possible to reproduce the passive elasticity of an intact myofibril. A similar reconstruction can also be done by numerically inverting the WLC model of polymer elasticity (Supplementary Information).

The physiological range of sarcomere lengths for a cardiac myofibril is 1.8–2.4 μm (ref. 29), corresponding to an extension range of 0–300 nm for I-band titin. The single-molecule data show that at an extension of 300 nm, the force reaches ~ 4 pN per I-band titin molecule. This force is about the same as that generated by a single myosin molecule⁴. At this force, the unfolding probability of the proximal tandem Ig region is low, $P_u = 0.1$. By contrast, the unfolding probability of the distal region is six orders of magnitude smaller. These results show that towards the end of the physiological range, unfolding of a few proximal Ig domains is possible whereas the distal domains always remain folded. If the unfolding probability of the proximal and distal Ig regions was zero, we would observe a purely entropic force–extension relationship (Fig. 4, black line). A purely entropic mechanism explains most of the extensibility of I-band titin in the physiological range, however, it departs significantly at higher extensions. These results suggest that unfolding of the proximal tandem Ig region may serve as a buffer to protect cardiac sarcomeres from developing damaging high forces. This becomes clear if we compare the effects of an over-extension to 450 nm. I-band titin will respond by unfolding several proximal Ig domains, limiting the force to ~ 7 pN. By contrast, if unfolding were

not possible, the force developed would exceed 40 pN per molecule, probably damaging sarcomeric structures. □

Methods

Protein engineering

All constructs were from human cardiac titin^{16,17}. Titin modules I4–I11, I4, I5, PEVK and N2B were cloned by polymerase chain reaction with reverse transcription (RT–PCR) from human heart poly(A)⁺ mRNA (Clontech) using the ThermoScript System (Gibco–BRL). Polypeptides I27₈, I28₈, I32₈, I34₈, I4₈, I5₈ and I27₃–N2B–I27₃ were constructed using a previously described method based on the identity of the sticky ends generated by *Bam*HI and *Bgl*II restriction enzymes^{18,19}, and then subcloned into pQE 80L (I4₈ and I5₈) or pQE 30 (I27₈, I28₈, I32₈, I34₈, I27₃–N2B–I27₃), (Qiagen). I27₁₂ was constructed using a non-palindromic *Ava*I restriction site (CTCGGG), as previously described¹⁸. (I27–PEVK)₃ was constructed using a similar method after *Eco*RI ligation of the two domains. (I27–PEVK)₃ and I4–I11 were cloned into pET–Ava I (ref. 18) while I27–I34 was cloned in pET 9d (ref. 11). The I27–I34 plasmid was a gift from M. Gautel¹¹. This protein has five changes to the sequence published for titin¹⁷: Thr 42 is replaced by Ala, and Ala 78 is replaced by Thr in the I27 module, Ala 53 is replaced by Thr in the I30 module, there is a deletion that includes the last two codons of I32 and 87 codons of the I33 domain^{18,30}, and a deletion of the Glu 89 codon of I34. The cloning strain was SURE-2 (Stratagene). The expression strains used were BL21 (DE3) (I27–I34), BLR (DE3) (I4₈, I5₈, (I27–PEVK)₃), BL21 (DE3) CodonPlus (I4–I11), SURE-2 (I27₈, I27₃–N2B–I27₃), and M15 (I28₈, I32₈, I34₈). Purification of recombinant proteins, from the soluble fraction of the bacterial lysate, was done by Ni²⁺-affinity chromatography in all the cases but for I4–I11, in which Co²⁺-affinity purification was used (Clontech). In the case of (I27–PEVK)₃ an additional size-exclusion fast performance liquid chromatography (FPLC) step was used. Proteins were kept at 4 °C in PBS with 5 mM dithiothreitol (DTT) and 0.2 mM EDTA, except for I27₈, I28₈, I32₈, I34₈, which were kept in 100 mM imidazole (pH 6.0). All the constructs used in this study have a His-tag at the amino terminus for affinity purification and two Cys residues at the carboxy terminus to promote covalent attachment of the protein to the gold-coated substrate.

AFM

Protein samples (3–10 μL , at a concentration of 10–100 $\mu\text{g ml}^{-1}$) were deposited onto freshly evaporated gold coverslips to allow the protein to adsorb onto the gold surface. Force–extension measurements were then carried out in PBS saline buffer (137 mM sodium chloride, 2.7 mM potassium chloride and 10 mM phosphate buffer, pH 7.4). The cantilevers are standard Si₃N₄ cantilevers from either Digital Instruments (with a typical spring constant of 100 mN m^{-1}) or TM Microscopes (with a typical spring constant of 12 mN m^{-1}). Every cantilever was calibrated in solution before use.

In situ recording of titin extensibility and force generation

The extensibility of the various I-band titin segments of rabbit cardiac myofibrils were measured using immunoelectron/immunofluorescence microscopy with a set of titin-specific antibodies⁶. Rabbit cardiac muscle expresses almost exclusively the N2B form¹⁶. Here, the technical names of the antibodies (T12, I17, I18, I20/22, MIR) were replaced for simplicity by consecutive numbers 1 to 5, with 1 being closest to the Z-disk and 5 being located at the A-band/I-band junction (Fig. 3). For each antibody type, the epitope-mobility data obtained over a range of sarcomere lengths (SLs) from 1.8 to 2.8 μm were pooled in SL bins of 50 nm. For each SL bin, the extension of a given titin segment was measured as the distance flanked by two nearest antibody epitopes: proximal Ig region, epitope 1 to epitope 2; N2B, 2 to 3; PEVK, 3 to 4; and distal Ig region, 4 to 5. The epitopes 3 and 4, which measure the extension of PEVK segment, include four additional Ig domains, hence the PEVK extension data were offset by 20 nm. Titin segment extension was then plotted against extension of the entire elastic titin in a half-sarcomere, $x(F)_{\text{I-band}}$ obtained as $x(F)_{\text{I-band}} = (\text{SL} - 1.8 \mu\text{m})/2$, to account for the functionally stiff titin in the sarcomere (1.6 μm in A-band, $2 \times 0.1 \mu\text{m}$ adjacent to Z-disk). The corresponding stretching force, F , was determined from mechanical recordings of the passive tension of isolated rabbit cardiac myofibrils^{6,27} immersed in a buffer solution (6 mM magnesium methanesulphonate, 5 mM dipotassium methanesulphonate, 4 mM Na₂ATP, 15 mM EGTA, with a total ionic strength of 200 mM adjusted with KOH in a 3-N-morpholino-propanesulphonic acid buffer, pH 7.1, 40 $\mu\text{g leupeptin ml}^{-1}$). Experimental protocols have been described²⁷. Passive force was recorded under quasi-steady-state conditions, that is, two to three minutes following a stretch to a new sarcomere length, to exclude viscous and viscoelastic force components that decay during stress relaxation.

Received 5 November 2001; accepted 14 June 2002; doi:10.1038/nature00938.

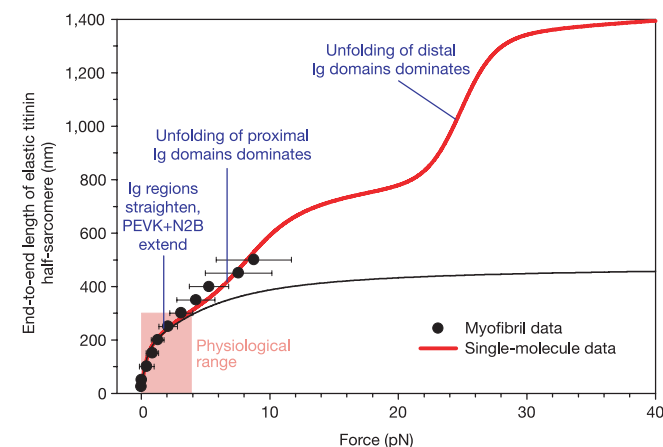


Figure 4 Single-molecule data predict the force–extension curve of cardiac muscle. The red line plots the calculated (equation (6)) end-to-end length of I-band titin versus a stretching force. The black line also plots equation (6), but in this case the unfolding probability of both the proximal and distal tandem Ig regions was set to zero. The symbols plot force–extension measurements from non-activated rabbit cardiac myofibrils. The values of the measured force were scaled to a single molecule assuming 6×10^9 titin molecules per mm^2 of cross-sectional area. The data show that the single-molecule data fully explain the force–extension relationship within and beyond the physiological range (coloured box).

1. Sigworth, F. J. & Neher, E. Single Na⁺ channel currents observed in cultured rat muscle cells. *Nature* **287**, 447–449 (1980).
2. Bustamante, C., Smith, S. B., Liphardt, J. & Smith, D. Single-molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol.* **10**, 279–285 (2000).
3. Smith, D. E. *et al.* The bacteriophage $\phi 29$ portal motor can package DNA against a large internal force. *Nature* **413**, 748–752 (2001).
4. Finer, J. T., Simmons, R. M. & Spudis, J. A. Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature* **368**, 113–119 (1994).
5. Lu, H. & Schulten, K. Steered molecular dynamics simulations of force-induced protein domain unfolding. *Protein Struct. Funct. Genet.* **35**, 453–463 (1999).
6. Linke, W. A. *et al.* I-band titin in cardiac muscle is a three-element molecular spring and is critical for maintaining thin filament structure. *J. Cell Biol.* **146**, 631–644 (1999).
7. Maruyama, K. Connectin/titin, giant elastic protein of muscle. *FASEB J.* **11**, 341–345 (1997).

8. Wang, K. Titin/connectin and nebulin: giant protein rulers of muscle structure and function. *Adv. Biophys.* **33**, 123–134 (1996).
9. Gregorio, C. C., Granzier, H., Sorimachi, H. & Labeit, S. Muscle assembly: a titanic achievement? *Curr. Opin. Cell Biol.* **11**, 18–25 (1999).
10. Trinick, J. & Tskhovrebova, L. Titin: a molecular control freak. *Trends Cell Biol.* **9**, 377–380 (1999).
11. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. & Gaub, H. E. Reversible unfolding of individual immunoglobulin domains by AFM. *Science* **276**, 1109–1112 (1997).
12. Kellermayer, M., Smith, S., Granzier, H. & Bustamante, C. Folding–unfolding transitions in single titin molecules characterized with laser tweezers. *Science* **276**, 1112–1116 (1997).
13. Tskhovrebova, L., Trinick, J., Sleep, J. A. & Simmons, R. M. Elasticity and unfolding of single molecules of the giant muscle protein titin. *Nature* **387**, 308–312 (1997).
14. Tskhovrebova, L. & Trinick, J. Direct visualization of extensibility in isolated titin molecules. *J. Mol. Biol.* **265**, 100–106 (1997).
15. Fisher, T. E., Marszalek, P. E. & Fernandez, J. M. Stretching single molecules into novel conformations using the atomic force microscope. *Nature Struct. Biol.* **7**, 719–724 (2000).
16. Freiburg, A. *et al.* Series of exon-skipping events in the elastic spring region of titin as the structural basis for myofibrillar elastic diversity. *Circ. Res.* **86**, 1114–1121 (2000).
17. Labeit, S. & Kolmerer, B. Titins, giant proteins in charge of muscle ultrastructure and elasticity. *Science* **270**, 293–296 (1995).
18. Carrion-Vazquez, M. *et al.* Mechanical and chemical unfolding of a single protein: a comparison. *Proc. Natl Acad. Sci. USA* **96**, 3694–3699 (1999).
19. Li, H. B., Oberhauser, A. F., Fowler, S. B., Clarke, J. & Fernandez, J. M. Atomic force microscopy reveals the mechanical design of a modular protein. *Proc. Natl Acad. Sci. USA* **97**, 6527–6531 (2000).
20. Li, H. B. *et al.* Multiple conformations of PEVK proteins detected by single-molecule techniques. *Proc. Natl Acad. Sci. USA* **98**, 10682–10686 (2001).
21. Best, R. B., Li, B., Steward, A., Daggett, V. & Clarke, J. Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation. *Biophys. J.* **81**, 2344–2356 (2001).
22. Marko, J. F. & Siggia, E. D. Stretching DNA. *Macromolecules* **28**, 8759–8770 (1995).
23. Bell, G. I. Models for the specific adhesion of cells to cells. *Science* **200**, 618–627 (1978).
24. Marszalek, P. E. *et al.* Mechanical unfolding intermediates in titin modules. *Nature* **402**, 100–103 (1999).
25. Trombitas, K., Freiburg, A., Centner, T., Labeit, S. & Granzier, H. Molecular dissection of N2B cardiac titin's extensibility. *Biophys. J.* **77**, 3189–3196 (1999).
26. Bueche, F. *Physical Properties of Polymers* 37 (Interscience, New York, 1962).
27. Linke, W. A. *et al.* Towards a molecular understanding of the elasticity of titin. *J. Mol. Biol.* **261**, 62–71 (1996).
28. Higuchi, H., Nakauchi, Y., Maruyama, K. & Fujime, S. Characterization of beta-connectin (titin 2) from striated muscle by dynamic light scattering. *Biophys. J.* **65**, 1906–1915 (1993).
29. Allen, D. G. & Kentish, J. C. The cellular basis of the length-tension relation in cardiac muscle. *J. Mol. Cell Cardiol.* **17**, 821–840 (1985).
30. Scott, K. A., Steward, A., Fowler, S. B. & Clarke, J. Titin; a multidomain protein that behaves as the sum of its parts. *J. Mol. Biol.* **315**, 819–829 (2002).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements

We thank H. Erickson for the electron microscope pictures of I27 polypeptides. W.A.L. thanks the German Research Foundation for a Heisenberg fellowship. This work was supported by the National Institutes of Health (J.M.F.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.M.F. (e-mail: jfernandez@columbia.edu).

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building

4 Crinan Street

London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silke Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.johnson@naturejp.com

Asia-Pacific Advertising Manager:

Kevyn Johnson

naturejobs

Inspiration and opportunity

African Americans have long been under-represented in US science, and statistics issued by the National Science Foundation have shown little progress over the years. The problem is self-perpetuating — without peers in science, minority students often lack inspiration. And without targeted funding programmes, lower-income minorities often lack opportunities. So any hint of progress is welcome. This summer, there have been some positive signs on both fronts.

Tshaka Cunningham, a graduate student at the Rockefeller University in New York, received encouragement from two sources. He recalls spending time as a child in a National Cancer Institute lab with his grandmother Alfreda Simmons, who, after gaining her degree from an all-black university, had difficulty pursuing graduate work. But although science intrigued him, Cunningham didn't see it as a viable long-term career.

So when he finished his degree at Princeton University, Cunningham took a job at drug firm Bristol-Myers Squibb, hoping to learn the business of drug discovery, and contemplating an MBA. But before matriculating, Cunningham talked to his former Princeton mentor Arnold Levine, then president of Rockefeller University. Levine convinced Cunningham to give science a go. Now, Cunningham is glad he stayed the course. He is now doing fascinating AIDS research and this summer received the David Rockefeller Fellowship.

And Cunningham isn't alone. This month, the US National Academy of Sciences awarded Ford Foundation Fellowships to 130 minority scholars, giving them the opportunity to pursue graduate degrees and postdoctoral work. These grants and Cunningham's success are positive — if small — steps in the face of a large problem, but only time will tell if they mark the reversal of a troubling long-term trend.

Paul Smaglik

Naturejobs editor



Contents

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS